25th International Conference on Computers in Education

"Technology and Innovation: Computer-based Educational Systems for the 21st Century"

Main Conference Proceedings

4 - 8 December 2017
Christchurch, New Zealand
MESSAGES

Message from the Conference Chair on our 25th Anniversary

On behalf of the organizing committee, I would like to extend my warm welcome to all delegates of the 25th International Conference on Computers in Education (ICCE) 2017, held for the second time in New Zealand. The year 2017 marks a milestone for ICCE as we celebrate our 25th anniversary. In celebrating our silver jubilee, we would like to express our sincerest appreciation to you, for your loyal and strong support through the years. When ICCE was first started as a biennial event in 1989, few would have imagined that it would grow and flourish beyond Taiwan where it had originated. Over the years, ICCE has received wide recognition as a highly reputable conference with participants coming from all four corners of the globe. Building on the continuous success of the conference series in recent years, the program aims to play the paramount role of connecting researchers in the Asia-Pacific region and beyond with international research communities to disseminate and share new ideas about the 21st century learning environments. This crucial role is aptly reflected in the conference theme of ICCE 2017 — “Technology and Innovation: Computer-based Educational Systems for 21st century learners”.

Four outstanding keynote speakers will share their insights across varying areas in the field of computers in education. They are 1). Mike Sharples from the Open University, UK who will highlight new directions in personalized learning from three different perspectives—open, informal and social; 2). Judy Kay from the University of Sydney, Australia who will present a vision for Personal Informatics for Learners (PIL) which builds on three foundations—i. Open Learner Models in Artificial Intelligence in Education, ii. Learning Analytics and Knowledge, iii. Personal Informatics; 3). Eric Klopfer from the Massachusetts Institute of Technology, USA who will draw upon his work in location-based Augmented Reality games, as well as work in Virtual Reality to highlight the design affordances of different mixes of reality for learning; 4). Tsukasa Hirashima from Hiroshima University, Japan who will talk about the problem-solving tasks where an additional characterization of problem-solving tasks by using two factors—i. well/ill-structured domain model and ii. well/ill-structured task setting will be highlighted.

In addition, there are three equally exciting theme-based talks — 1). “Large-scale teacher professional development for effective technology integration” by Sahana Murthy from the Indian Institute of Technology Bombay, India; 2). “Effective TELL for self-regulated learning and collaborative learning” by Yoshiko Goda from Kumamoto University, Japan; 3). “Fostering collective creativity and 21st century literacies through CSCL and formative learning analytics environments” by Jennifer Pei-Ling Tan from the National Institute of Education, Singapore.
These seven presentations surely capture the essence of the aforesaid conference theme and will act as catalysts to transform the learning environments and inspire us to re-think the roles of technological tools, pedagogical strategies and innovative theories in the 21st century.

Indeed, organizing such a large-scale conference requires the concerted efforts and unwavering support from the conference organizing committee members and conference paper reviewers. To recognize these amazing individuals, their names are listed in the proceedings. I would also like to take this opportunity to record my sincerest appreciation to all the kind individuals who have rendered their help in every possible way to make this conference a reality. I am also grateful to all the paper authors and registered participants for their exciting academic contributions to the fruitful intellectual exchange in this conference. I hope all delegates will not only have opportunities to renew friendships, forge new friendships and professional collaborations but also able to reflect upon and celebrate with us more than a quarter century of our history. I hope that you will have a productive and fun-filled time at this very special conference and leave Christchurch—a beautiful and picturesque city—with warm and everlasting memories.

Thank you!

Su Luan WONG
Conference Chair
(Malaysia)
Message from Local Organizing Committee

Kia ora! Welcome to New Zealand!

This is the first time an ICCE conference is hosted on New Zealand’s South Island. Early December is the start of summer in New Zealand, and weather is usually perfect; we do hope that will be the case in 2017! We are, however, sure that all participants will enjoy the Rydges Latimer hotel, well known for excellent food and service. The conference venue overlooks the scenic Latimer Square, an iconic green space surrounded by lush trees. The hotel is in the city centre, and the participants will have an opportunity to see Christchurch being rebuilt after the devastating earthquakes in 2010/2011. The city is re-emerging as one of New Zealand’s most vibrant and exciting cities. Lonely Planet listed it as one of the top 10 cities to visit in 2013.

There are many possibilities for sightseeing before and after the conference. Christchurch is the gateway to New Zealand’s South Island. You can ski, bungy jump, hike, mountain bike, raft, surf, swim, golf, see whales, dolphins and seals, visit wineries and gardens, music events, shop, be entertained and awed, and so much more, all within 2 hours of Christchurch.

New Zealand, as a society, strongly values education and innovation in educational practices. In recent years, due to extensive efforts of academics from the University of Canterbury and other universities, Computer Science has been taught in high schools. Many innovative examples of using computers have been developed during this initiative and related work – hence our conference theme.

It is a great privilege to be able to host the 2017 conference and share our beautiful country. We thank the APSCE Executive Committee for giving us this wonderful opportunity. Our sincere thanks to all the sponsors, the standing committee, the International Program Committee, reviewers, authors, participants and student volunteers. We trust all of you will enjoy the conference, and take home great memories from Aotearoa New Zealand.

Ngā mihi

LOC Chair
Tanja Mitrovic

LOC Co-Chair
Moffat Mathews
Message from the International Program Coordination Chairs

The International Conference on Computers in Education (ICCE) is an annual conference series encompassing a broad range of issues related to using Information and Communication Technology (ICT) for education, organized by the Asia-Pacific Society for Computers in Education (APSCE). ICCE 2017 is taking place in Christchurch, New Zealand, from December 4th to 8th 2017. It aims to bring together researchers from all over the world to share and exchange research and to develop and deploy new ideas that span the field of Computers in Education. Following the tradition of previous conferences in this series, ICCE 2017 is organized as a meta-conference where there are seven sub-conferences focusing on specialized themes. Each sub-conference is organized by a program committee appointed by the respective Special Interest Group (SIG – see http://www.apsce.net/sigs_list.php?id=1026).

These sub-conferences are:
C1: ICCE Conference on Artificial Intelligence in Education/Intelligent Tutoring System (AIED/ITS) and Adaptive Learning (AL)
C2: ICCE Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences (LS)
C3: ICCE Conference on Advanced Learning Technologies (ALT), Learning Analytics and Digital Infrastructure
C4: ICCE Conference on Classroom, Ubiquitous, and Mobile Technologies Enhanced Learning (CUMTEL)
C5: ICCE Conference on Digital Game and Digital Toy Enhanced Learning and Society (GTEL&S)
C6: ICCE Conference on Technology Enhanced Language Learning (TELL)
C7: ICCE Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)

The International Program Committee consists of a strong and dedicated team that includes the Conference Chair, the Program Coordination Chair and Co-Chair, seven executive Sub-Conference Chairs and 343 experts in the field of Computers in Education from 39 different countries or economies. Former ICCE local organizing and program coordination chairs have played the role of consultants in overseeing the conference organization process.

The conference received a total of 213 papers (135 full, 50 short, and 28 posters) from 30 different countries or economies. Table 1 provides the submissions by country of the authors in individual papers. All papers were subjected to a rigorous review process by 3-5 reviewers from the respective Sub-Conference program committees. After the reviews were completed, a meta-review was provided for each paper. 660 reviews and meta-reviews were received in total. After a discussion period within the individual program committees led by the Sub-Conference Executive Co-Chairs and Co-Chairs, recommendations were made to the Coordination Committee Chair and Co-Chair, who oversee the review process and quality for all sub-conferences. This resulted in 49 full, 78 short, and 62 poster acceptances across all of the 7 sub-conferences. The overall acceptance rate for full papers is 23%. The acceptance rate for the full papers in the individual Sub-Conference closely mirrored the overall acceptance rate. This is a testimony to the continued maintenance of the quality of presentations in our conference. The complete statistics of paper acceptance is shown in Table 2.
In addition to full papers, short papers, posters, ICCE2017 include program components such as keynote speeches, theme based invited speeches, workshops, tutorials, panels, Work-in-Progress Posters (WIPP), Extended Summary (ES), Doctoral Student Consortia (DSC), and Early Career Workshop (ECW). All the papers in these program components are published in separate proceedings with their own ISBN numbers. Pre-conference events are held on the first two days of the conference, including workshops, Doctoral Student Consortia and tutorials.

Table 1: Distribution of Paper Submissions for ICCE 2017

<table>
<thead>
<tr>
<th>Country of authors</th>
<th>Number of submissions</th>
<th>Country of authors</th>
<th>Number of submissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>145</td>
<td>Korea</td>
<td>3</td>
</tr>
<tr>
<td>China</td>
<td>74</td>
<td>Canada</td>
<td>4</td>
</tr>
<tr>
<td>Taiwan</td>
<td>67</td>
<td>Switzerland</td>
<td>3</td>
</tr>
<tr>
<td>New Zealand</td>
<td>25</td>
<td>Finland</td>
<td>2</td>
</tr>
<tr>
<td>Brazil</td>
<td>25</td>
<td>Indonesia</td>
<td>2</td>
</tr>
<tr>
<td>Singapore</td>
<td>22</td>
<td>Norway</td>
<td>2</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>22</td>
<td>Sweden</td>
<td>2</td>
</tr>
<tr>
<td>Philippines</td>
<td>20</td>
<td>Ecuador</td>
<td>2</td>
</tr>
<tr>
<td>India</td>
<td>19</td>
<td>Sudan</td>
<td>2</td>
</tr>
<tr>
<td>Thailand</td>
<td>19</td>
<td>UK</td>
<td>1</td>
</tr>
<tr>
<td>USA</td>
<td>18</td>
<td>France</td>
<td>1</td>
</tr>
<tr>
<td>Australia</td>
<td>17</td>
<td>Kenya</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>5</td>
<td>Macau</td>
<td>1</td>
</tr>
<tr>
<td>Malaysia</td>
<td>5</td>
<td>Spain</td>
<td>1</td>
</tr>
<tr>
<td>Tunisia</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Breakdown of submission and acceptance by sub-conference for ICCE 2017

<table>
<thead>
<tr>
<th>Sub-conference</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>21 (5)</td>
<td>24 (7)</td>
<td>31 (11)</td>
<td>26 (11)</td>
<td>10 (5)</td>
<td>12 (5)</td>
<td>11 (5)</td>
<td>135 (49)</td>
</tr>
<tr>
<td>Short</td>
<td>4 (8)</td>
<td>5 (8)</td>
<td>13 (22)</td>
<td>12 (21)</td>
<td>4 (5)</td>
<td>3 (7)</td>
<td>9 (7)</td>
<td>50 (78)</td>
</tr>
<tr>
<td>Poster</td>
<td>1 (9)</td>
<td>1 (8)</td>
<td>2 (11)</td>
<td>5 (10)</td>
<td>7 (9)</td>
<td>7 (7)</td>
<td>5 (8)</td>
<td>28 (62)</td>
</tr>
<tr>
<td>Total</td>
<td>26 (22)</td>
<td>30 (23)</td>
<td>46 (44)</td>
<td>43 (42)</td>
<td>21 (19)</td>
<td>22 (19)</td>
<td>25 (20)</td>
<td>213 (189)</td>
</tr>
</tbody>
</table>

We would like to thank all who have contributed to making ICCE2017 a successful conference. First of all, we would like to thank all paper authors for your contributions and for choosing ICCE 2017 as the outlet to present your research. We would also like to thank the IPC Executive Chairs and members who undertook the responsibility of reviewing and selecting papers representing research of high quality. Specially thanks to our keynote and invited speakers for accepting our invitations and bring inspiring research to ICCE2017 participants. The Local Organization Committee deserves a big thank you for their hard work under the tremendous time pressure.

We hope that all participants will find the activities in ICCE2017 interesting and inspiring, and have opportunities to meet old friends and establish new professional collaborations. Furthermore, we hope that participants will enjoy not only the academic activities, but also the beautiful scenery and vibrant and exciting culture experiences in Christchurch.
International Program Coordination Chairs:

Wenli CHEN (Chair),
National Institute of Education,
Nanyang Technological University
Singapore

Jie Chi YANG (Co-Chair),
National Central University
Taiwan

Ahmad Fauzi Bin Mohd AYUB (Co-Chair),
Universiti Putra Malaysia
2 ORGANIZATION

ORGANIZED BY:
Asia Pacific Society for Computers in Education

HOSTED BY:
University of Canterbury, New Zealand

CONFERENCE CHAIR:
Su Luan WONG, Universiti Putra Malaysia, Malaysia

INTERNATIONAL PROGRAM COMMITTEE (IPC)

IPC Coordination Chair:

Wenli CHEN, Nanyang Technological University, Singapore

Co-Chairs:
Jie-Chi YANG, National Central University, Taiwan
Ahmad Fauzi Bin Mohd Ayub, Universiti Putra Malaysia, Malaysia

Workshop/Tutorial/Interactive Event Coordination Chair:

Yusuke HAYASHI, Hiroshima University, Japan

Co-Chairs:
Moffat MATHEWS, University of Canterbury, New Zealand
Thepchai SUPNITHI, Nectec, Thailand

Poster Coordination Chair:

Chengjiu YIN, Kobe University, Japan

Co-Chairs:
Jihong DING, Zhejiang University of Technology, China
Sagaya AMALATHAS, UNITAR International University, Malaysia

Doctoral Student Consortia Chair:

Fu-Yun YU, National Cheng Kung University, Taiwan

Co-Chair:
Sahana MURTHY, Indian Institute of Technology Bombay, India

Members:
Ben CHANG, National Central University, Taiwan
Maiga CHANG, Athabasca University, Canada
Gautam BISWAS, Vanderbilt University, USA
Jon MASON, Charles Darwin University, Australia
Early Career Workshop Chair:
Lung-Hsiang WONG, Nanyang Technological University, Singapore

Co-Chairs: Hyo-Jeong SO, Ewha Woman's University, Korea
Maiga CHANG, Athabasca University, Canada

Panel Chair:
Morris S.Y. JONG, Chinese University of Hong Kong

Co-Chair: Maiga CHANG, Athabasca University, Canada

Extended Summary Coordination Chair:
Jie Chi Yang, National Central University, Taiwan

Merit Scholarship Committee Chair:
Mas NIDA, Universiti Putra Malaysia, Malaysia

Co-Chair: Weiqin CHEN, Oslo and Akershus University College of Applied Sciences, University of Bergen, Norway

Theme-based Conference Program Co-Chairs:

C1: ICCE Conference on Artificial Intelligence in Education/Intelligent Tutoring System (AIED/ITS) and Adaptive Learning (AL)
Gautam BISWAS, Vanderbilt University, USA (Executive chair)
Sergey SOSNOVSKY, German Center for Artificial Intelligence, Germany
Amali WEERASINGHE, the University of Adelaide, Australia
Tsukasa HIRASHIMA, Hiroshima University, Japan

C2: ICCE Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences (LS)
Pratim SENGUPTA, University of Calgary, Canada (Executive chair)
Beaumie KIM, University of Calgary, Canada
Kate THOMPSON, Griffith University, Australia
David HUANG Junsong, Nanyang Technological University, Singapore
Corey BRADY, Vanderbilt University, USA

C3: ICCE Conference on Advanced Learning Technologies (ALT), Learning Analytics and Digital Infrastructure
Jon MASON, Charles Darwin University, Australia (Executive chair)
Eunice SARI, UX Indonesia
Yong KIM, Korea National Open University, Korea
Judith AZCARRAGA, De La Salle University, Philippines
C4: **ICCE Conference on Classroom, Ubiquitous and Mobile Technologies Enhanced Learning (CUMTEL)**
Yanjie SONG, Hong Kong Institute of Education, Hong Kong (Executive chair)
Nelson BALOIAN, Universidad de Chile, Chile
Ben CHANG, National Central University, Taiwan
Su CAI, Beijing Normal University, China
Chengjiu YIN, Kobe University, Japan

C5: **ICCE Conference on Digital Game and Digital Toy Enhanced Learning and Society (GTEL&S)**
Hiroyuki MITSUHARA, University of Tokushima, Japan (Executive chair)
Morris S. Y. JONG, The Chinese University of Hong Kong, Hong Kong
Gwo-Jen HWANG, National Taiwan University of Science and Technology, Taiwan
Fernández-Manjón BALTASAR, Universidad Complutense de Madrid, Spain
Chris HOLDEN, University of New Mexico, United States

C6: **ICCE Conference on Technology Enhanced Language Learning (TELL)**
Ting-Chia HSU, National Taiwan Normal University, Taiwan (Executive chair)
Agnes KUKULSKA-HULME, The Open University of UK, UK
Yun WEN, Singapore Centre for Chinese Language, Nanyang Technological University, Singapore
Jiyou JIA, Peking University, China
Greg KESSLER, Ohio University, USA
Yoshiko GODA, Kumamoto University, Japan

C7: **ICCE Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)**
Daner SUN, Hong Kong Institute of Education, Hong Kong (Executive chair)
Mi Song KIM, University of Western Ontario, Canada
Jingyun WANG, Kyushu University, Japan
Elizabeth Ruilin KOH, Nanyang Technological University, Singapore

**Consultants:**
Lung Hsiang WONG, Nanyang Technological University, Singapore
Weiqin CHEN, Oslo and Akershus University College of Applied Sciences, University of Bergen, Norway
Bo JIANG, Zhejiang University of Technology, China
Sahana MURTHY, Indian Institute of Technology Bombay, India
Sridhar IYER, Indian Institute of Technology Bombay, India
Special Interest Groups (SIG) Chairs:

AIED/ITS/LS: Yusuke HAYASHI, Hiroshima University, Japan
CSCL/LS: Seng Chee TAN, Nanyang Technological University, Singapore
ALT/OC/S: Tore HOEL, Oslo and Akershus University College of Applied Sciences, Norway
CUMTEL: Xiaoqing GU, Eastern China Normal University, China
GTel&S: Maiga CHANG, Athabasca University, Canada
TELL: Masanori YAMADA, Kyushu University, Japan
PTP: Hiroaki OGATA, Kyushu University, Japan
DICTAP: Md Khambari MAS NIDA, University Putra Malaysia, Malaysia

Educational Use of Problems: Tsukasa HIRASHIMA, Hiroshima University, Japan
LA&EDM: Weiqin CHEN, Oslo and Akershus University College of Applied Sciences, University of Bergen, Norway

LOCAL ORGANIZING COMMITTEE (LOC)

Chairs: Antonija MITROVIC
Co-Chair: Moffat MATHEWS
Sponsorship Chair: Hossein SARRAFZADEH
Publicity Chair: Amali WEERASINGHE
Publicity Chair: Nilufar BAGHAEI
Webmaster: Jay HOLLAND
Student Volunteer Coordinator: Geela FABIC
Proceedings Coordinator: Ashish SHARMA
Graphic Design: Day Design Creative
C1 PC members

Ryan BAKER, University of Pennsylvania, USA
Ben DANIEL, University of Otago, New Zealand
Agneta GULZ, Linkoping and Lund Universities, Sweden
Tomoko KOJIRI, Kansai University, Japan
Nguyen-THINH LE, Humboldt-Universität Berlin, Germany
Elvira POPESCU, University of Craiova, Romania
Olga SANTOS, UNED, Spain
Merlin SUAREZ, De La Salle University, Philippines
Carsten ULLRICH, DKFI, Germany
Michelle P. BANAWAN, Ateneo de Davao University, Philippines
Krittaya BUNCHONGCHIT, Mahidol University, Bangkok, Thailand
Mario CHACON, Tecnológico de Costa Rica
Jose Luis Perez De La CRUZ, Universidad de Malaga, Spain
Ben Du BOULAY, Informatics Department, University of Sussex, UK
Claude FRASSON, University of Montreal, Canada
Shinobu HASEGAWA, Japan Advanced Institute of Science and Technology, Japan
Yusuke HAYASHI, Hiroshima University, Japan
Bastiaan HEEREN, Open University of the Netherlands
Sharon HSIAO, Arizona State University, USA
Tomoya HORIZU, Kobe University, Japan
Sridhar IYER, Indian Institute of Technology Mumbai
Akihiro KASHIHARA, Univ. of Electro-Communications, Japan
Chad LANE, University of Illinois, USA
Lydia LAU, University of Leeds, UK
Chi-Jen LIN, Fo Guang University, Taiwan
Noboru MATSUDA, Texas A&M University, USA
Tatsunori MATSUI, Waseda University, Japan
Moffat MATHEWS, University of Canterbury, Christchurch, New Zealand
Tanja MITROVIC, University of Canterbury, Christchurch, New Zealand
Kazuhisa MIWA, Nagoya University, Japan
Riichiro MIZOGUCHI, Japan Advanced Institute of Science and Technology, Japan
Jaclyn OCUMPAUGH, University of Pennsylvania, USA
Helen PAIN, University of Edinburgh, UK
Ma. Mercedes T. RODRIGO, Ateneo de Manila University, Philippines
Sagaya SABESTINAL, UNITAR International University, Malaysia
M. SASIKUMAR, CDAC, India
Natalia Stash, Eindhoven University of Technology, Netherlands
Kazuhisa SETA, Osaka Prefecture University, Japan
Thepchai SUPNITHI, National Electronics and Computer Technology Center, Thailand

C2 PC members
Amanda DICKES, Harvard University, USA
Jun OSHIMA, Shizuoka University, Japan
Peter REINMANN, University of Sydney, Australia
Chee Kit LOOI, Nanyang Technological University, Singapore
Chen-Chung LIU, National Central University, Taiwan
Ronghuai HUANG, Beijing Normal University, China
Xiaoqing GU, East China Normal University, China
Heisawon JEONG, Hallym University, Republic of Korea
Yanyan LI, Beijing Normal University, China
Rose LIANG, National University of Singapore, Singapore
Joyce Hwee Ling KOH, Nanyang Technological University, Singapore
Seng Chee TAN, Nanyang Technological University, Singapore
Timothy Khe Foon HEW, The University of Hong Kong, Hong Kong
Huang-Yao HONG, National Chengchi University, Taiwan
Gwendoline Choon Lang QUEK, Nanyang Technological University, Singapore
Jianhua ZHAO, South China Normal University, China
Qiyun WANG, Nanyang Technological University, Singapore
Doris CHOO, Nanyang Technological University, Singapore
Yun WEN, The Swiss Federal Institute of Technology, Switzerland
Madhuri Mavinkurve, Indian Institute of Technology Bombay, India
Mrinal Patwardhan, Indian Institute of Technology Bombay, India
Sanjay Chandrasekharan, Tata Institute of Fundamental Research, India
Yotam Hod, University of Haifa, Israel
Shiang-Kwei Wang, New York Institute of Technology, USA
Johanna Pöysä-Tarhonen, Finnish Institute for Educational Research, Finland
Youngju Lee, Korea National University of Education, Korea
Dani Ben-Zvi, University of Haifa, Israel
Daniel Bodemer, University Duisburg-Essen, Germany
Rosemary Luckin, The London Knowledge Lab, UK
Jan-Willem Strijbos, University of Groningen, the Netherlands
Moon-Heum Cho, Sungkyunkwan University, Korea
Wenli Chen, Nanyang Technological University, Singapore
Sahana Murthy, Indian Institute of Technology Bombay, India
Suneeta Pathak, Communication Analytics Pte. Ltd., Singapore
C3 PC members

Luis ANIDO RIFON, University of Vigo
Leigh BLACKALL, RMIT University
Sandy BRITAIN, New Zealand Ministry of Education
Irene CHEN, National Changhua University of Education
Yong-Sang CHO, Korea Education and Research Information Service
Jean-Noel COLIN, University of Namur
Giovanni FULANTELLI, Istituto per le Tecnologie Didattiche
Dai GRIFFITHS, University of Bolton
Tore HOEL, Oslo and Akershus University College of Applied Sciences
Ulrich HOPPE, University of Duisburg-Essen, Germany
Yih-Ruey JUANG, Jinwen University of Science and Technology
Khalid KHAN, Charles Darwin University
Tatsuhiro KONISHI, Shizuoka University
Lam-for KWOK, City University of Hong Kong
Jaeho LEE, University of Seoul
Chien-Sing LEE, Universiti Tunku Abdul Rahman
Jerry LEESON, University of Adelaide
Dobrila LOPEZ, Eastern Institute of Technology, New Zealand
Roberto MARTINEZ-MALDONADO, University of Technology Sydney
Kenji MATSUURA, Tokushima University
Kiyoshi NAKABAYASHI, Chiba Institute of Technology
Kourosh NESHATIAN, University of Canterbury, New Zealand
Kuo-Liang OU, National Hsinchu University of Education
Manuel Caeiro RODRIGUEZ, University of Vigo
Hitoshi SASAKI, Takushoku University
Jin Gon SHON, Korea National Open University
Marcus SPECHT, Open University of the Netherlands
Riina VUORIKARI, Institute for Prospective Technological Studies, European Commission
Maggie M. WANG, The University of Hong Kong
Tsuneo YAMADA, The Open University of Japan

C4 PC members

Lars BOLLEN, University of Twente, Netherlands
Peijing CAO, National Institute of Education Sciences, China
Chih-Hung CHEN, Graduate from National Taiwan University of Science and Technology
Tzung-Shi CHEN, National University of Tainan, Taiwan
Guang CHEN, Beijing Normal University, China
JunMing CHEN, National Museum of Natural Science, Taiwan
Kwok Shing CHENG, The Education University of Hong Kong
Feng-Kuang CHIANG, Beijing Normal University, China
Tosti CHIANG, National Taiwan Normal University, Taiwan
Chih-Ming CHU, National Ilan University, Taiwan
Hui-Chun CHU, Soochow University, Taiwan
Daniel CHURCHILL, The University of Hong Kong
Sabrina EIMLER, University of Ruhr-West, Germany
Haiguang FANG, Capital Normal University, China
Chaohua GONG, Southwest University, China
Naka GOTODA, Kagawa University, Japan
Ping HE, Tianjin University, China
Ting-Chia HSU, National Taiwan Normal University, Taiwan
Gwo-Jen HWANG, National Taiwan University of Science and Technology, Taiwan
Marc JANSEN, University of Ruhr-West, Germany
Jianli JIAO, South China Normal University, China
Bo JIANG, Zhejiang University of Technology, China
Yih-Ruey JUANG, Jinwen University of Science and Technology, Taiwan
Chorng-Shiuh KOONG, National Taichung University, Taiwan
Mouri KOUSUKE, Tokyo University of Agriculture and Technology, Japan
Nguyen-Thinh LE, Humboldt-Universität zu Berlin, Germany
Ping LI, The Chinese University of Hong Kong, China
Chen-Chung LIU, National Central University, Taiwan
Jun LIU, Capital Normal University, China
Gi-Zen LIU, National Cheng Kung University, Taiwan
Chiu-Pin LIN, National Tsing Hua University, Taiwan
Sven MANSKE, University of Duisburg, Germany
Yuichi ONO, Tsukuba University, Japan
Junjie SHANG, Peking University, China
Jun-Ming SU, National University of Tainan, Taiwan
Daner SUN, The Education University of Hong Kong
Hanyu SUNG, Chung Yuan Christian University, Taiwan
Yancy TOH, Nanyang Technological University, Singapore
Noriko UOSAKI, Osaka University, Japan
Sally WAN, The Chinese University of Hong Kong
C5 PC members
Liz BACON, University of Greenwich Old Royal Naval College, UK
Francesco BELLOTTI, University of Genoa, Italy
Ben CHANG, National Central University, Taiwan
Maiga CHANG, Athabasca University, Canada
Yuan-Jen CHANG, Central Taiwan University of Science and Technology, Taiwan
Nian-Shing Chen, National Sun Yat-Sen University, Taiwan
Zhi-Hong CHEN, National Taiwan Normal University, Taiwan
Hercy CHENG, Central China Normal University, China
Sridhar CHIMALAKONDA, IIIT-Hyderabad, India
Tsung-Yen CHUANG, National University of Tainan, Taiwan
Muhammet DEMIRBILEK, Suleyman Demirel University, Turkey
Larbi ESMAVI, Athabasca University, Canada
Mathew Joseph GAYDOS, SUNY Korea, Korea
Susan GWEE, The English Language Institute of Singapore
Toshihiro HAYASHI, Kagawa University, Japan
Robert HELLER, Athabasca University, Canada
Huei-Tse HOU, National Taiwan University of Science and Technology, Taiwan
Jun HU, Eindhoven University of Technology, Netherlands
Chung-Yuan HSU, National Pingtung University, Taiwan
Mingfong JAN, National Central University, Taiwan
Venkatesh V. KAMAT, Goa University, India
Beaumie KIM, University of Calgary, Canada
Kinshuk, University of North Texas, United States
Siu Cheung KONG, The Education University of Hong Kong, Hong Kong
Paul KUO, National Palace Museum, Taiwan
Rita KUO, New Mexico Institute of Mining and Technology, USA
Jungmin KWON, Seoul National University, Korea
Fong-Lok LEE, The Chinese University of Hong Kong, Hong Kong
Tsai-Yen LI, National Chengchi University, Taiwan
Hao-Chiang Koong LIN, National University of Tainan, Taiwan
Chang-Yen LIAO, Central China Normal University, Taiwan
Rossitza MARINOVA, Concordia University of Edmonton, Canada
Saurabh MEHTA, VIT, Mumbai, India
Motoki MIURA, Kyushu Institute of Technology, Japan
Wolfgang MUELLER, University of Education Weingarten, Germany
Kuo-Liang OU, National HsinChu University of Education, Taiwan
Michal PTASZYNSKI, Kitami Institute of Technology, Japan
Ma. Mercedes RODRIGO, Ateneo de Manila University, Philippines
Demetrios G. SAMPSON, Curtin University, Australia
Chun-Yi SHEN, Tamkang University, Taiwan
Juling SHIH, National University of Tainan, Taiwan
Masanori SUGIMOTO, Hokkaido University, Japan
Kaoru SUMI, Future University Hakodate, Japan
Chuen-Tsai SUN, National Chiao Tung University, Taiwan
Shu-Yuan TAO, Takming University of Science and Technology, Taiwan
Ming Hsin TSAI, Asia University, Taiwan
Hongxue Harris WANG, Athabasca University, Canada
Wing-Kwong WONG, National Yunlin University of Science and Technology, Taiwan
Yang YAN, Changchun Normal University, China
Hsi-Hsun YANG, National Yunlin University of Science and Technology, Taiwan

C6 PC members
Rebecca BABCOCK, University of Texas of the Permian Basin, USA
Mei Mei CHANG, National Pingtung University of Science and Technology, Taiwan
Sherry Ying-hsueh CHENG, National Pin-Tung University of Science and Technology, Taiwan
Jozef COLPAERT, University of Antwerp, Belgium
Liliana CUESTA, Universidad de La Sabana, Colombia
Additional reviewers
Wan-Chen CHANG, National Central University, Taiwan
Kae NAKAYA, University of Tokyo, Japan
Yen-Ning SU, National Cheng Kung University, Taiwan
Chen Hsieh, JUN, National Central University, Taiwan

C7 PC members
Alain BREULEUX, McGill University, Canada
Andrea VALENTE, University of Southern, Denmark
Azilawati JAMALUDIN, Nanyang Technological University, Singapore
Brendan FLANAGAN, Kyoto University, Japan
Chih-Kai CHANG, National University of Tainan, Taiwan
Cho Young HOAN, Seoul National University, Korea
Daisuke KIMURA, Global Incubation x Fostering Talents (GiFT), Japan
Emanuela MARCHETTI, Lillebaelt Academy University of Applied Sciences, Denmark
Fleur R. PRINSEN, Hogeschool Rotterdam, Netherlands
Huiying CAI, Jiangnan University, China
Ivica BOTICKI, University of Zagreb, Croatia
Kate THOMPSON, Griffith University, Australia
Katerina GEORGOULI, Technological Educational Institute of Athens, Greece
Kometani YUSUKE, Waseda University, Japan
Kosuke KANEKO, Kyushu University, Japan
Gary K.W. WONG, The University of Hong Kong, Hong Kong
Guillaume SCHILTZ, ETH Zurich, Switzerland
Gyeong Mi HEO, McGill University, Canada
Jane GICHURU, University of Western Ontario, Canada
Jason LEE, Nanyang Technological University, Singapore
Jennifer MASTERS, University of Tasmania, Australia
Jennifer Tan PEI-LING, Nanyang Technological University, Singapore
Jinbao ZHANG, Beijing Normal University, China
Jing LENG, East China Normal University, China
Johanna POYSA-TARHONEN, University of Jyvaskyla, Finland
Jun-Ming SU, National University of Tainan, Taiwan
Lara RUFFOLO, Independent scholar
Lee Shu SHING, Nanyang Technological University, Singapore
Longkai WU, Nanyang Technological University, Singapore
Ping C LIESER, Washington University, US
Lynde TAN, Western Sydney University, Australia
Ma LUO, East China Normal University, China
Michael HOOVER, The University of British Columbia, Canada
Mutlu CUKUROVA, University College London, UK
Rhoda Myra Garces BACSAL, Nanyang Technological University, Singapore
Pearl CHEN, California State University, USA
Pei-Shan TSAI, National Taipei University of Technology, Taiwan
Peter SEOW, Nanyang Technological University, Singapore
Pey-Tee OON, University of Macau, Macau
Ping LI, The Education University of Hong Kong, Hong Kong
Qing MA, The Education University of Hong Kong, Hong Kong
Qiyun WANG, Nanyang Technological University, Singapore
Romee LEE, University of British Columbia, Canada
Roya GHOLAMI, University of Illinois at Springfield, USA
Lay Hoon SEAH, Nanyang Technological University, Singapore
Sevil AKAYGÜN, Boğaziçi University, Turkey
Silin WEI, Hangzhou Normal University, China
Siu-yung JONG, The Chinese University of Hong Kong, Hong Kong
Chew Lee TEO, Ministry of Education, Singapore
Vitaliy POPOV, University of San Diego, USA
Xinghua WANG, Nanyang Technological University, Singapore
Yam San CHEE, Nanyang Technological University, Singapore
Yanjie SONG, The Education University of Hong Kong, Hong Kong
Yao XIONG, University of Pennsylvania, USA
Yiu Chi LAI, The Education University of Hong Kong, Hong Kong
Yuichi ONO, Tsukuba University, Japan
Yun WEN, Singapore Centre of Chinese Language, Singapore
Zhihong WAN, The Education University of Hong Kong, Hong Kong
3 Distinguished Researcher and Early Career Researcher Awards

2015
APSCE Distinguished Researcher Award
• Lung-Hsiang Wong, Nanyang Technological University

APSCE Early Career Researcher Award
• Morris Siu-yung Jong, The Chinese University of Hong Kong

2014
APSCE Distinguished Researcher Award
• Hiroaki Ogata, Kyushu University, Japan

2011
APSCE Distinguished Researcher Award
• Antonija Mitrovic, University of Canterbury, New Zealand
• Chen-Chung Liu, National Central University, Taiwan

APSCE Young Researcher Leader Award
• Ben Chang, National Chiayi University, Taiwan
• Wenli Chen, Nanyang Technological University, Singapore

2009
APSCE Distinguished Researcher Award
• Fu-Yun Yu, National Cheng Kung University, Taiwan
• Tsukasa Hirashima, Hiroshima University, Japan

APSCE Young Researcher Leader Award
• Hiroaki Ogata, University of Tokushima, Japan
Distinguished Researcher and Early Career Researcher Awards

APSCE conducts the schemes of Distinguished Researcher Award (DRA) and Early Career Research Award (ECRA) every year. DRA recognizes an active APSCE Member who has showed distinguished academic accomplishments and contributions in the field of Computers in Education. The awardee of DRA must be under 50 at the time of nomination. ECRA recognizes an active APSCE Member in the early stages of his/her career no later than 10 years after receipt of the doctoral degree who has produced international quality research outputs, and be able to demonstrate aspirations and potentials to achieve world-leading status. These awards also recognize contributions to APSCE (e.g. related activities to ICCE or SIGs) or APSCE appointments (e.g. Executive Committee member of APSCE). These awards were launched in 2009 for provision every two years. Starting from 2014, the awards are provided every year. Each awardee is required to contribute a paper to RPTEL, the journal of APSCE, within two years of receiving his/her award.

In 2017, APSCE received nil nomination for DRA and two nominations for ECRA. After the qualification check by the Award Subcommittee, the two ECRA nominees were accepted as candidates for the award. The first voting phase for selecting the ECRA finalist was subsequently carried out, and one candidate was accepted to proceed to the final selection phase. The ECRA winner was decided in the final selection phase through a competitive winner-voting exercise: the ECRA finalist has to receive at least “2/3” votes of agreement from all APSCE EC members for the award-winning.

We are now pleased to announce the award winner of 2017 APSCE ECRA: Dr. Jon Mason.

In the following page, you will find a record of Dr. Mason’s scholarship and his services to APSCE. The ECRA award winner will receive an APSCE ECRA Certificate, monetary reward, and will have his name publicized on APSCE web site and on the ICCE proceedings. He will also receive nomination for the Theme-based Invited Speaker for ICCE in a future year.

APSCE Award Subcommittee:
Siu Cheung KONG (Chair),
Su Luan WONG,
Fu-Yun YU
2017 APSCE Early Career Researcher Award Winner

Dr. Jon Mason
Charles Darwin University, Australia

Dr. Jon Mason is a senior lecturer within the School of Education at Charles Darwin University (CDU) in Australia where he leads research into digital education futures. He also holds adjunct positions as a Professor within the Department of e-Learning at Korea National Open University, a Professor of Educational Technology at East China Normal University, and as a Project Consultant for the Advanced Innovation Center for Future Education at Beijing Normal University.

Jon joined CDU in 2012 initially as the Director of e-Learning for the Centre for School Leadership, Learning and Development. At this time, he was still completing his PhD with Queensland University of Technology under the supervision of Professor Hitendra Pillay (Education) and Professor Christine Bruce (Information Systems). Jon had attempted PhD studies earlier while working at the University of Melbourne but found the demands of full-time work too much at the time.

Jon's PhD dissertation was titled The Why Dimension - Opening Frontiers for Digital Learning. Set against a background narrative of the evolution of the application of digital technologies in education, this study identified that technology-based scaffolding for questioning is a field ready for innovation. The role of computers in education has come a long way from the days of processing 'if-then' queries and is now enabled by countless innovations that support knowledge sharing, openness, flexibility, and independent inquiry. Given the diversity of inputs into the field of technology, enhanced learning the investigation was necessarily interdisciplinary – drawing from information science, educational theory, developmental psychology, educational technology, and knowledge management. A pivotal construct in the thesis identifies five activities associated with the word why. Together, these activities are defined as the ‘why dimension’ – asking, learning, understanding, knowing, and explaining why. Findings highlight deficiencies and biases in mainstream search-based approaches to inquiry, which default to privilege the retrieval of information as distinct from explanation. Despite their power, search engine technologies abbreviate questioning and decompose inquiry. Instrumental to human sense making, the ‘why dimension’ provides a conceptual framework for addressing this shortcoming through the development of ‘sense-making technologies’.

Jon's recent entrance into academia followed over a decade of working at the nexus of
government services, education, and international standardization. Prior to this, he worked in IT management at the University of Melbourne and before that as a private music teacher.

Jon has an extensive publication record and has served in numerous editorial roles for international projects, journals and books. These include serving as Editor for the International e-Framework for Education and Research, a collaboration involving government agencies from the UK, Netherlands, Australia, and New Zealand. He is currently the Lead Editor of a thematic series for Research and Practice in Technology Enhanced Learning (RPTEL) focused on ‘Positioning for Digital Learning Futures’. Jon has been an elected member of the Executive Committee of the Asia Pacific Society for Computers in Education since 2012. His current research agenda spans the learning sciences, digital learning futures, sense making, question technologies, student questioning skills, and wisdom in education.
4 KEYNOTE AND INVITED SESSIONS

KEYNOTE SPEAKER 1

*New directions in personalized learning: Open, informal, social*

Mike Sharples
Open University, UK

Abstract

A goal of educational technology since the 1930s has been to adapt teaching to the personal needs of each student. Significant developments have included programmed instruction, branched instruction, intelligent tutoring systems, and adaptive courseware. Personalized learning is coming back into prominence with the development of new techniques for linking learning analytics to adaptive teaching. Research challenges include how to enable personalization of informal and inquiry-led learning, and how to link personalization with learning through conversation and social networking.

Personalized open learning must offer opportunities for students from widely differing backgrounds to learn in ways that match their needs and abilities. This requires new designs for flexibility of timing, pace, facilitation and assessment. For informal learning, personalization must align with changes in context, learning materials co-created by students, and self-directed study. In social networked learning, students need support to merge their individual pathways through the curriculum into shared goals, positive interdependence and productive conversation.

I shall discuss recent work at The Open University on predictive analytics and flexible pathways for learners, as part of a strategic university initiative in personalized open learning. Our iSpot and nQuire-it platforms combine informal science learning with personalization through reputation.
management and student authoring. Adaptive crowdsourcing may offer mechanisms for personalized social networked learning.
Abstract
This talk presents a vision for Personal Informatics for Learners (PIL). This builds on three foundations. The first, and oldest of these, comes from the decades of research on Open Learner Models (OLMs) in Artificial Intelligence in Education. These fields strive to create high quality teaching systems with personalisation driven by a fine-grained, carefully crafted model of the learner. The second is the far newer, but very fast growing LAK, Learning Analytics and Knowledge. This arose with wide deployment of learning technology that captures huge quantities of learning data and the tantalising possibilities such data offers. The third, Personal Informatics (PI) comes from Ubicomp research, aiming to harness personal sensor data about diverse aspects of life, particularly for health and wellness. PIL is grounded on the view that the learner should be empowered to take responsibility for their own learning. To achieve this, PIL needs to make progress on many fronts. We need mechanisms for collecting the right learning data and managing it effectively. We also need new classes of the interfaces and systems that enable learners to control their data, do personal data mining and engage in meta-cognitive processes of reflection, self-monitoring, planning.

The talk begins by reviewing the broad scope and vision described above. Taking that lens, I then present a two sets of case studies. The first relates to group-work, with interfaces onto data harvested as groups use an online collaboration tool or collaborate around an interactive tabletop or wall display, with diverse models, from the simple to richer ones built by data mining. Key insights come from the series of studies, both in the lab and in authentic classrooms. The second set of case studies is for individual learning, ranging from mastering computer software and curriculum-wide learner modelling to personal informatics, for health and wellness, harnessing data from activity trackers, virtual reality and mobile food logging. The talk concludes with key lessons learnt and a research agenda for PIL.
From Augmented to Virtual Learning: Design Affordances of Different Mixes of Reality for Learning

Eric Klopfer
Massachusetts Institute of Technology, USA

Abstract
Mixed realities that combine digital and real experiences are now becoming a true reality. These experiences are being delivered over smartphones as well as increasingly accessible and practical head mounted displays. This ubiquity of devices is in turn making mixed reality the next digital frontier in entertainment, education and the workplace. But what do we know about where these technologies have value? Where do they add to the learning experience? And what theories and evidence can we generate and build upon to provide a foundation for using these technologies productively for learning? We have been working on mixed realities in education for over a decade and have started to learn about where, when and for whom they can add value. Part of this understanding stems from differentiating the wide variety of mixed realities and focusing on affordances. Landscape based Augmented Realities, popularized by Pokemon Go, have fundamentally different affordances than smartphone based Virtual Realities like Google Cardboard, which in turn are different than immersive experiences delivered by headsets like the Oculus Rift and HTC Vive. The core of our work has been doing research and development to identify these affordances that match with key learning challenges, particular in Science, Technology, Engineering and Mathematics (STEM). In this talk, I will draw upon our work in location-based Augmented Reality games, as well as work in Virtual Reality. In the realm of Augmented Reality, I will discuss a long series of design experiments through which we have learned about where these technologies play an important role in learning, primarily around socio-scientific issues. In the space of Virtual Reality our newest designs and experiments focus on the concept of scale, and how we can use Virtual Realities to teach about STEM systems at radically different scales. This talk will provide a history and overview of these experiences, including iterations of design research experiments.
Tsukasa Hirashima
Hiroshima University, Japan

Abstract
Problem-solving tasks are often categorized into two types: ill-structured one and well-structured one, in the context of education/learning, cognitive science and artificial intelligence. Then, from an educational viewpoint, ill-structured tasks are further more important because they are useful to promote a learner to think about target learning contents deeply and to master computational or logical thinking skills, including metacognition and self-regulation. This talk presents an additional characterization of problem-solving tasks by using two factors, (1) well/ill-structured domain model and (2) well/ill-structured task setting. Here, a well-structured domain model provides a problem space as a set of states and a set of operators linking one state with the next. Then, a well-structured task setting provides a specific problem space with an initial state and a goal state. Based on this characterization, “moderately ill-structured tasks” can be defined as a category of tasks specified by “well-structured domain model” and “ill-structured task setting”. If a task is set in well-structured domain model, it is possible to realize computer-based monitoring and diagnosis of learner’s activities for the task. If the task setting is ill-structured, for example, open-ended, a learner is required to engage in the task as ill-structured one. Therefore, moderately ill-structured tasks are promising to realize computer-based intelligent support for solving the tasks, while keeping educational advantages of ill-structured tasks. In this paper, a definition of moderately ill-structured tasks is described. As a method to realize scaffolding and intelligent support, information structure open approach where the domain model is represented as information structure and open for learners to direct manipulation are proposed. In this talk, using arithmetic word problems as an example of learning contents, (1) well-structured domain model of arithmetic word problems, and (2) design of moderately ill-structured task as “problem-posing assignment” based on the information structure model are introduced. Moreover, (3) implementation and practical uses of several intelligent learning environments that support learners to solve the moderately ill-structured tasks as problem-posing are reported.
Large-scale teacher professional development for effective technology integration

Sahana Murthy
Indian Institute of Technology Bombay, India

Abstract
As the use of digital technology in teaching and learning proliferates, teachers face challenges of how to choose appropriate tools and integrate them in their teaching. Teachers need explicit training in creating student-centred learning designs that harness the affordances of the technology, and in implementing them effectively in their classroom context. Several teacher professional development programs exist, but two key challenges are how to sustain results beyond the duration of the program, and how to scale such efforts.

In this talk, I will describe how we address the sustainability and scalability issues in teacher professional development for learner-centered technology integration practices. As part of Project TUET (Teacher Use of Educational Technology), a flagship project in the department of Educational Technology at IIT Bombay, we have developed models, training program designs and tools that empower teachers in effective technology integration practices within their own teaching-learning context. We have developed the A2I2 (Attain-Align-Integrate-Investigate) model and its associated design principles of immersivity, pertinency and transfer of ownership. We have implemented A2I2-based teacher professional development programs in face-to-face, blended and fully online modes both at the school level and higher education level. We have scaled up such programs up to 4000 participants via learner-centered MOOCs. I will present evidence of how teachers’ design expertise has evolved, and how some teachers have successfully transferred ownership of the problem of effective technology integration via action research in their own classroom. I will discuss our efforts towards building a community of practice over the past five years, and conclude with recommendations for designing and implementing large-scale and potentially sustainable teacher professional development programs.
Yoshiko Goda
Kumamoto University, Japan

Abstract
In this presentation, I will discuss ways to merge technology with effective and active language learning through the applications of Self-Regulated Learning (SRL) and Collaborative Learning (CL), focusing on learning support. There are four main parts to the presentation: (1) SRL support for course completion and dropout reduction and the limitations of SRL, (2) CL support to increase engagement, (3) bridging SRL and CL for quality interactions, and (4) current issues and future implications.

How do people behave and regulate their learning in TELL? When online learning materials were assigned with a due date, there were found to be seven learning types: (1) procrastination, (2) learning habit, (3) random, (4) diminished drive, (5) early bird, (6) chevron, and (7) catch-up. When the relationships between the learning types and their learning outcomes were analyzed, the results showed that the students with the learning habit type scored significantly higher on the test than did students with the procrastination type. These results imply that regulated learning could increase learning effectiveness and lead to better learning outcomes in e-learning. However, the problem is that most of us (generally 70 to 80 percent) are said to be procrastinators. Therefore, it is difficult to conclude that all procrastinators learn ineffectively. Then, the concepts of active and passive procrastination will be introduced.

The communicative approach has been encouraged for effective and active language learning. Not only SRL but CL is also essential to increase the size, depth, and fluency of language proficiency. To increase quality interactions among students, Community of Inquiry (CoI) was employed for quality interactions in our research projects, with increments of social, cognitive, and teaching presences. The design and support for effective CL and bridging SRL and CL will be presented through demos of our developed systems.

Finally, limitations of SRL, current issues with CL, and the combination of SRL and CL will be discussed. Innovation and creative skills are required as the 21st century skills. These skills should be applied to language learning, as well. The agenda for innovative TELL will be proposed for our future development.
Effective TELL for Self-Regulated Learning and Collaborative Learning

Jennifer Pei-Ling Tan
National Institute of Education, Singapore

Abstract

Collaborative, creative and critical literacies are essential to young people’s productive participation in 21st century lifeworlds. Yet, research has shown that the classroom can often be an uncomfortable and problematic space for developing such skills and dispositions. One major challenge lies in how we can more effectively assess and scaffold the development of these ‘new(er) literacies’ in learners, at both individual and collective levels, and as they occur naturalistically in peer interactions during acts of learning. With more appreciation for the dynamic and non-linear nature of 21st century literacies and their constitutive socio-interactional processes, educators worldwide are increasingly cognizant of the limitations of conventional assessment and pedagogic modalities. To this end, well-designed computer-supported collaborative learning (CSCL) environments that leverage on formative social learning analytics (LA) may bring new affordances to bear on this educational imperative of our time.

In this talk, I will showcase one techno-pedagogical innovation that exemplifies how the purposeful coupling of CSCL and formative LA affordances can serve to enhance collective creativity and critical literacies in secondary students within the disciplinary domain of English language learning. Alongside the explication of key design principles and empirical learning gains, I will also foreground the pedagogical dilemmas and challenges encountered throughout the iterative cycles of design, enactment, adoption and diffusion. In doing so, I hope to underscore both the educational promises and problems that arise as the ‘rubber’ of well-intentioned learning innovations ‘hits the road’ of entrenched socio-institutional beliefs and practices in mainstream schooling.
PANEL 1  *Meta-Synthesis and Trends Reports of Educational Technologies and Pedagogies: What is their Value Proposition?*

**ABSTRACT**  
The need for nurturing the global citizens and innovators of the 21st century has prompted the education field to re-examine what it takes to deliver or facilitate quality teaching and learning. Thus, annual synthesis efforts have ensued where experts are convened to analyse trends and produce reports to help practitioners, policymakers, researchers and even learners to explore both the bigger picture and specific aspects of techno-pedagogical diversity. Such informed synthesis is typically harder to access through regular means. Two such initiatives are the Horizon Project and Innovating Pedagogy. While both report series have been rather well-exposed among their target audience over the years, they are not meant to be final products or echo chambers; instead, they serve to catalyse collective reflections and exchanges within the community in order to rise above and challenge the viewpoints put forward by the reports. This panel aims to provide a platform for eliciting opinions and critiques on the purposes, the synthesis methods, the value and the limitations of such or similar report series. Key personnel behind the two stated report series will provide insights on how the reports were developed; other panelists will offer their critical views on the reports from the perspective of general "users". Through the panel interactions, it is hoped that the participants will be able to establish a renewed understanding on how to make meaningful interpretation and use of the report series to impact the directions of research in and practice of technology-enhanced learning.

**FACILITATOR**  
Lung-Hsiang WONG, Nanyang Technological University (NTU), Singapore

**PANELISTS**  
Chee-Kit LOOI, Nanyang Technological University (NTU), Singapore

Mike SHARPLES, The Open University, UK

Yuan GAO, Smart Learning Institute, Beijing Normal University (BNU), China

Jon MASON, Charles Darwin University (CDU), Australia
Meta-Synthesis and Trends Reports of Educational Technologies and Pedagogies: What is their Value Proposition?

Lung-Hsiang WONG*a, Mike SHARPLESb, Yuan GAOc, Jon MASONd & Chee-Kit LOOIt

aLearning Sciences Lab, National Institute of Education, Nanyang Technological University, Singapore
bThe Open University, UK
cSmart Learning Institute, Beijing Normal University, China
dCharles Darwin University, Australia
*lhwong.acad@gmail.com

Abstract: The need for nurturing the global citizens and innovators of the 21st century has prompted the education field to re-examine what it takes to deliver or facilitate quality teaching and learning. Thus, annual synthesis efforts have ensued where experts are convened to analyse trends and produce reports to help practitioners, policymakers, researchers and even learners to explore both the bigger picture and specific aspects of techno-pedagogical diversity. Such informed synthesis is typically harder to access through regular means. Two such initiatives are the Horizon Project and Innovating Pedagogy. While both report series have been rather well-exposed among their target audience over the years, they are not meant to be final products or echo chambers; instead, they serve to catalyse collective reflections and exchanges within the community in order to rise above and challenge the viewpoints put forward by the reports. This panel aims to provide a platform for eliciting opinions and critiques on the purposes, the synthesis methods, the value and the limitations of such or similar report series. Key personnel behind the two stated report series will provide insights on how the reports were developed; other panelists will offer their critical views on the reports from the perspective of general "users". Through the panel interactions, it is hoped that the participants will be able to establish a renewed understanding on how to make meaningful interpretation and use of the report series to impact the directions of research in and practice of technology-enhanced learning.

Keywords: Learning technology landscape, pedagogical innovations, synthesis reports, Horizon Reports, Innovating Pedagogy Reports

1. Introduction

Educators are increasingly needing to respond to ‘21st century’ agendas that focus on nurturing global citizens through innovative approaches to teaching and learning. This requires us to re-examine our practices and what is essential to facilitating quality teaching and learning. While the waves of continuous technological advancement provide opportunities they also challenge existing norms in almost all fields. Consequently, educators, policymakers and scholars may feel overwhelmed at times due to the immense variety and relentlessness of changes of the learning technology landscape. To ensure we do due diligence, we need to identify novel, spinoff or hybrid pedagogical approaches that continue to challenge conventional beliefs about teaching and learning, as well as technological advances that can influence pedagogies. It is no surprise, then, that efforts to identify, distil or synthesize the key trends have also become institutionalised. Two of such initiatives are Horizon Project and Innovating Pedagogy. Both initiatives analyse trends and produce reports based upon expert consultation and both initiatives aim to guide practitioners, policymakers, researchers and learners to explore aspects of techno-pedagogical diversity that might be otherwise more difficult to understand.

Initiated by the New Media Consortium (NMC) in 2002, the Horizon Project has been publishing annual reports (with multiple specialised editions in each year) that chart emerging technologies for teaching, learning and assessment (typically six types of technologies introduced in
each report). From 2014 onwards, the reports go beyond a technology-focused positioning by incorporating the explication of the key trends accelerating technology adoption in education institutions and challenges that impede such adoption.

In contrast, the Innovating Pedagogy Report (Sharples et al, 2016) series places its emphasis on exploring ten forms of pedagogy in each year that are already in currency but have not yet have a profound influence on education. Some of the surveyed pedagogical approaches may not require technological support, yet there is always a potential for technologies to play a part in enhancing the learning experience and outcomes. The report series is curated by a group of educational researchers from The Open University in the United Kingdom, led by Professor Mike Sharples. The series was established in 2012. In 2015 and 2016, academics from SRI International in Stanford, US, and National Institute of Education, Nanyang Technological University, Singapore were respectively invited to join the Open University counterparts in contributing to the writing efforts.

While both report series are rather well-exposed among their target audience over the years, the interesting questions that we would like to ask relate to how members of the ICCE community should treat the reports. In what ways could we optimise the value that the reports provide, particularly for scholars and practitioners of our region? While the report series represent informed commentary about the field they can also catalyse collective reflections and exchanges within the community in order to rise above and challenge the viewpoints put forward by the reports.

This panel aims to provide a platform for eliciting opinions and critiques on the purposes, the synthesis methods, the value and the limitations of such or similar report series. Through the panel interactions, it is hoped that the participants will be able to establish a renewed understanding on how to make meaningful interpretation and use of the report series to impact the directions of research in and practice of technology-enhanced learning.

2. Abstracts of Individual Panelists’ Presentations

2.1 Interpreting Pedagogy-led Innovations for Education
(Mike SHARPLES)

The Innovating Pedagogy reports are intended for teachers, policy makers, academics and anyone interested in how education may change over the next ten years. By ‘innovative pedagogies’ we mean theories and practices of teaching, learning and assessment for the modern, technology-enabled world. Whereas other report series cover the influence of technologies on education, this explores the influences and affordances of pedagogy. Arguably, the greatest innovations in education over the past 40 years have been led by pedagogy not technology – such as widespread adoption of collaborative learning in schools and colleges, and problem-based learning in workplaces and universities.

Each Innovating Pedagogy report has been written by a small group of academics from the Institute of Educational Technology at The Open University, with the two most recent reports produced in collaboration with leading international institutes in learning and technology. We have shared ideas, proposed innovations, read research papers and blogs, and commented on each other’s draft contributions. We compile the report by first producing a long list of new educational terms, theories, and practices, then reducing these to ones that have the potential to provoke major shifts in educational practice. Each report introduces ten pedagogies that are starting to influence educational practice or offer opportunities for the future.

In our reports, we aim to understand and acknowledge learning in a world of interactive digital technologies. A focus on technologies could run the risk of chasing each invention up and down the switchback of innovation, marketing, hype and obsolescence. Some devices for education that have long been out of fashion, such as adaptive teaching machines, language labs, and integrated learning systems, reappear in new guises for the next wave of technology. By examining innovative pedagogies, we aim to ride this roller coaster of technology adoption, highlighting methods of teaching, learning and assessing that can be successful now and into the future.
In the panel, how the Chinese Horizon Reports especially focusing on Chinese educational contexts were developed as well as their values are presented. The Chinese Horizon Reports used an improved Delphi-Method to collect data, including five phases on the procedure: 1. Establishment of expert committee, in which 65 panel experts for K-12 education and 85 panel experts for higher education from 8 areas all over the China (east China, south China, central China, north China, north east, west east, and Hong Kong, Macao, and Taiwan regions) were included. 2. Desktop research, in which the experts were asked to review materials relevant to the research questions including news, reports and essays for emerging technologies on the wiki to complement their prior knowledge related to the current research. 3. Answering research questions, in which expert committee discussed and answered the four main research questions on the wiki that relate to the key trends accelerating technology adoption, significant challenges impeding technology adoption, and important developments in educational technology. 4. Online voting, in which each panel expert was required to rank the emerging technologies that will be adopted in next five years as well as the key trends and the significant challenges in adopting them; meanwhile, the important development, trends, and challenges need to be classified into different time and difficulty dimension (1yrs, 2-3yrs, 4-5yrs for important development, short-term, middle-term, and long-term for key trends, solvable, difficult, and wicked for challenges). 5. Project example collection, in which the practical examples related to the important development of emerging technologies, key trends, and significant challenges were collected from all over the China to show how they are exactly applied in teaching and learning of China.

The Chinese Horizon Reports have significant values for different groups of readers. For the policy-makers, the reports help them to understand the development trends of educational technology so that they can make and adjust educational policies and plans accordingly. For the university and school leaders, the reports provide them with forward-looking data which introduce new and emerging technologies for universities and schools to improve the teaching quality. For the EdTech companies and enterprises, the reports help them to grasp the direction of development for educational technology as well as to put more resources into the development of emerging technologies listed in the report. The data of the reports also open a door for scholars, researchers, and students majoring in educational technology, facilitating them to do substantial research to explore how to use these emerging technologies in education positively and productively. In addition, the project examples collected in the reports showed how to combine emerging technologies and instruction suitably, which help the teachers and educators improve the use of emerging technologies in their own classes. As for media and journalists, the Chinese Horizon Report helps them to learn about the dynamic states of the ICT in education in a timely manner.

2.3 Re-Positioning Sense-Making in Technology Enhanced Learning

What characterizes the expanding field of educational technology? Churn, constant change, waves of digital revolution, innovation, disruption, opportunity, choice, and many other things. Because of such dynamics it is a field replete with buzzwords, cute new terminology, acronyms, reports, monographs, predictions, and hype. If you’ve ever heard of Cisco’s CEO John Chambers you’ve probably also heard that he once said the e-learning market was poised to make the invention of email look like a ‘rounding error’ (Horton, 2001, p. 24). Chambers said this when the tech sector was experiencing enormous growth as the spearhead of the so-called new economy; but, he also said this just prior to the dotcom bust. In time, Chambers may be proved right but the lessons from history provide useful perspective about the nature of cycles and change. Foresight is useful when accurate; however, it is usually hindsight that balances the picture.

Given that the innovations driving the Internet are enabling while also being disruptive it is not hard to find a value proposition for big picture commentaries of the dynamics shaping change in educational technology. In the broader field of information technology Gartner has been routinely producing iterations of its ‘hype cycle’ for two decades (Mullany, 2016). At its core, the hype cycle
depicts the transition from hype to maturity and adoption of new technologies. While not scientifically valid, this representation makes lots of sense in the corporate world of decision-making. Likewise, for most of us in the field of educational technology, we need to make sense of all that is going on, whether it is the likely impact of innovation in technology or a consequent shift in pedagogical practice. Making sense of change is a natural human activity, and in the words of Brenda Dervin, is a ‘mandate of the human condition’ (2005, p. 27). More recently, Madsbjerg (2017) provides a compelling case for the role of the humanities in ‘sensemaking’ in the contemporary era increasingly driven by big data and algorithms. Thus, it is of interest to the theme being examined by this panel that ‘sense-making’ was picked up for a while by the Horizon Reports between 2010-2012 but has not really had much prominence since. It has also appeared in the 2015 Innovating Pedagogy report (Sharpley, et al., 2015, p.28). Arguably, it is terminology that needs to remain part of our conversations because it is fundamental to understanding and learning (Mason, 2014) And, for too long, constructivist theory has not adequately addressed the role of sense-making. In the case of connectivism, the topic is addressed in a tentative way by Downes (2012). This also raises an important point – useful synthesis and big picture commentaries also appear in the form of posts by prominent bloggers such as Stephen Downes.

As an educator focused on what digital technologies make possible both report series are invaluable resources. Apart from providing some temporary order to a volatile picture these reports at times offer practical advice. Of course, like any academic writing, these reports can be criticized for misplaced emphasis or some other shortcoming. But this is what distinguishes useful academic commentary – it can be scrutinized and tested.

Having said that, it will be of interest to learn what both reports might have to say in the future on the topic of data literacy. This topic is quite distinct from the what has generally been understood as digital literacy and needs to be handled differently. It is not just about being competent with using computers and the internet, it is about understanding the nature of data, its origins and destinations, the black box algorithms that produce it, knowing how to examine data visualizations, and how to discern the fake from the real in an era of post-truth. This topic also aligns closely with sense-making and calls into question the adequacy of those competencies commonly described as 21st century skills (Mason, Khan, & Smith, 2016). A key question to emerge in recent times concerns privacy. As Tim Berners-Lee recently said when reflecting on 28 years after he first invented the web ‘we’ve lost control of our personal data’ (Berners-Lee, 2017). Such hindsight provides cautionary advice as we continue to recalibrate our relationship with technology and education systems. Thus, how do we make sense of the consequences of innovation and how can we fine-tune our foresight?

### 3. Discussion and Conclusion

Indeed, regular and timely reports of trends of pedagogies or technologies as described in the earlier sections serve multiple purposes – for policy makers to know about broad and potential trends, practitioners to know what works for them, scientists to know the answer to questions or issues that have been well studied, and researchers to know what is known about a field and what are likely gaps in our current knowledge. Such trend reports are produced through revised versions or Delphi method or tapping on a panel of experts. Another type of reports are the meta-meta-synthesis or meta-synthesis reports. These use statistical tools for combining findings from different studies with the goal of identifying patterns that can inform practice. For example, Visible Learning (2009) by Hattie is based on more than 800 meta-analyses and more than 50,000 studies involving 250 million students. At time of writing of Visible Learning for Math, Grades K-12 (Hattie, 2017), the database included 1200 meta-analyses with more than 70,000 studies and 300 million students. A meta-synthesis synthesizes what is currently known about a given topic and can result in strong recommendations about the impact and effect of a specific practice. There is also an approach called qualitative meta-synthesis which attempts to integrate results from a number of different but inter-related qualitative studies.

While the Innovating Pedagogy report has an interpretative intent towards examining innovative pedagogies, the Chinese Horizon Reports has a stronger aggregating intent, as with meta-meta-analysis studies such as those of Hattie. Mason argues for the sense-making intent. With much commonalities...
and yet enough diversity, we hope the panel will provide a stimulating discussion, centred on addressing these questions:

• As a community, what do we see as the role of these trends and meta-meta-synthesis reports? Should ICCE promote the role of these meta-synthesis and trends reports? What are the forums to publish, share, discuss and promote these reports and studies?

• Should we as researchers lead, support or be involved in producing more of these reports? If anything, meta-synthesis reports are helpful for young and beginning researchers in any given area, and thus they can help shape the direction of research in the Asia-Pacific region.

• What are the incentives and structures to produce more meta-synthesis and trends reports as compared with specific studies? The fact that such publications tend to elicit substantial number of citations should be a good incentive for mid-tier or senior researchers to embark on such studies.

References


ABSTRACT There is a number of emerging educational needs planned to be integrated into school curriculum. Computational thinking education, an important trend in school education in the 21st century, is one of these educational needs advocated to be integrated into school curriculum. The integration of all these educational trends into school curriculum at a time may have two problems, although it seems to address the concerns of all relevant stakeholder groups. The first problem is a lack of sufficient curriculum time to incorporate all newly added curriculum elements. The second problem is a lack of coherence in implementing the initiatives as there are so many stakeholders involved in such integration and there are so many concerns about the curriculum. To respond to the accommodation of all changes incurred in the emergence of digitalization in every aspect of daily life, there is a need to make a holistic review of the existing school systems for well-planned changes related to the school buildings, school structure, school curriculum and pedagogy. This panel aims to discuss critical issues for the need of making a holistic review of the existing school systems to accommodate the need of such changes incurred in the digitization in every aspect of the daily life in the digital era and how to transform.

FACILITATOR Siu Cheung KONG, The Education University of Hong Kong, Hong Kong

PANELISTS
Chee-Kit LOOI, Nanyang Technological University (NTU), Singapore
Mike SHARPLES, The Open University, UK
Yuan GAO, Smart Learning Institute, Beijing Normal University (BNU), China
Jon MASON, Charles Darwin University (CDU), Australia
A Holistic Review of School Systems for Accommodating Changes Incurred in the Digital Era: Is It a Must?

Siu Cheung KONG a,*, Tak-Wai CHAN b, Chee Kit LOOI c, Gautam BISWAS d, & Nian-Shing CHEN e

a The Education University of Hong Kong, Hong Kong
b National Central University, Taiwan
c Nanyang Technological University, Singapore
d Vanderbilt University, USA
e National Sun Yat-sen University, Taiwan

Abstract: There is a number of emerging educational needs planned to be integrated into school curriculum. Computational thinking education, an important trend in school education in the 21st century, is one of these educational needs advocated to be integrated into school curriculum. The integration of all these educational trends into school curriculum at a time may have two problems, although it seems to address the concerns of all relevant stakeholder groups. The first problem is a lack of sufficient curriculum time to incorporate all newly added curriculum elements. The second problem is a lack of coherence in implementing the initiatives as there are so many stakeholders involved in such integration and there are so many concerns about the curriculum. To respond to the accommodation of all changes incurred in the emergence of digitalization in every aspect of daily life, there is a need to make a holistic review of the existing school systems for well-planned changes related to the school buildings, school structure, school curriculum and pedagogy. This panel aims to discuss critical issues for the need of making a holistic review of the existing school systems to accommodate the need of such changes incurred in the digitization in every aspect of the daily life in the digital era and how to transform.

Keywords: Changes, comprehensive holistic review, computational thinking education, digital era, school curriculum

1. Introduction

Learners in the digital era are expected to be interest-driven creators, who are capable of becoming lifelong creators through the immersion in the interest-driven learning process (Chan et al., 2015; Chen et al., 2015). The technological support in the e-Learning process provides school leaders and teachers in school education around the world with opportunities to engage young learners in sustained interest-driven learning activities which foster learners to develop interests in learning, capabilities in creation, and habits of learning (Chan et al., 2015; Wong et al., 2015). The design and implementation of curriculum framework for computational thinking development is one of the promising approaches for the school education sectors to nurture interest-driven creators, in which the learning activities expose learners to see the world through a computational lens from the problem-solving experiences and practices arising in programming (Kong, 2016).

This panel aims to discuss critical issues for the need of making a holistic review of the existing school systems to accommodate the need of such changes incurred in the digitization in every aspect of the daily life in the digital era and how to transform. The panel is expected to provide participants
with insights related to preparing school learners living in the digital era to accommodate changes incurred in the emergence of digitalization in every aspect of daily life.

2. Abstracts of Individual Panelists’ Presentation

2.1 Forces That Drives Changes in Asian Education
(Tak-Wai CHAN)

Technological changes occur at an exponential rate while human beings are far much slower in adopting the changes. Actually, schools in the whole world are undergoing a transformation through rapid cycles of changes and adoptions. In Asia, there are some particularly strong forces that drive schools to change. For example, the current dominating examination-driven education in Asia cannot meet the future need of our next generations. In this panel, I would like to identify some of these forces and to describe how we scholars in this field can play a vital role in gearing the course of this transformation by developing theory and building compelling and impactful examples for our society to follow.

2.2 Holistic Review: What Does It Entail?
(Chee Kit LOOI)

Holistic reviews of national curricula in the light of the changing landscape of education needs and drivers including technology advances would seem to be a sensible political and educational imperative. Some of the pertinent questions are: How soon do these waves of changes accumulate before necessitating the review and the re-think? When do one recognize and problematize the coming wave is a critical one that needs forward planning and addressing, and not one that can be addressed as it comes along (for example, let others take the lead, and then learn from their experiences)? What are the trade-offs in changing the eco-system of education to address the new changes via a holistic review, versus incremental tweaking or ground-up experimentations in changing the status quo?

Taking the example of computational thinking (CT), many countries and regions recognize the importance of it as a basic literacy, and have already implemented or are deliberating policy changes to integrate CT or coding into the school systems. For CT to be instituted as the formal education, policy decisions need to be designed, tweaked or refined at a more systemic level that entails a holistic review of the curricula including strategic decisions on whether CT is an addition to the formal curriculum or it replaces something else in the curriculum or is integrated into existing subjects. A holistic review would also include decisions that would consider the appropriate grade level to introduce CT and the forms of assessing the development of CT competencies.

2.3 How Should School Systems Accommodate to Changes in the Digital Era
(Gautam BISWAS)

The advent and advances in the digital era and the information technology revolution are making transformative changes in our daily lives and the way we go about doing business. Given this scenario, it seems natural that the education in schools must change to accommodate these changes and provide students with 21st century and lifelong learning skills. In abstract terms, learning processes must provide mechanisms by which students can acquire and interpret knowledge, organize this knowledge in a way that allows application to problem solving processes, and the ability to verify that the knowledge acquired is correct, and consistent with past knowledge. Learning should also accommodate transfer in a variety of forms. In my presentation, I will discuss how technology can be introduced in the classroom to support many of these processes.
2.4 Challenges Faced by Different Stakeholders while Implementing ICT in Classrooms (Nian-Shing CHEN)

In implementing new technologies into a school environment, different stakeholders are involved, from the generating of the idea till the implementation of the technology into the classroom. Every stakeholder would face with different challenges and issues. In this panel, I will try to elaborate these issues raised by different stakeholders in accordance to learners, teachers/educators, parents, school administrators, developers and researchers. I will also discuss about how to evaluating these challenges and issues before conducting an actual implementation.

3. Discussion and Conclusion

There is an increasing attention to the importance of computational thinking for the next generation to creatively and effectively solve daily life problems in the digital era. This creates an emerging need for school education sectors to develop young learners with computational thinking in order to succeed in the digital era. This further drives a growing trend of designing and implementing curriculum framework for computational thinking development in school education around the world. The compact schedule of formal school education curriculum already covers an array of important subject learning contents. The panel provides an insightful discussion about a holistic review of the existing school systems to prepare school leaders and teachers for a strategic promotion of computational thinking development in view of the compact school curriculum.

The panel first consists of panellist presentations which share the important issues arguing for the need of a holistic review of the existing school systems in school education across different countries / regions around the world. The panel then discusses the most desirable school systems including the design of school buildings, school structures, school curriculum content and pedagogy to achieve the goals of desirable school systems fitting the need of learners living in the digital world. Panellists may share appropriate approaches to tackling these issues from the perspectives of their countries / regions. There may be common issues across countries / regions and there may be unique issues in different countries / regions. The discussion focuses on the need to make a holistic review of the design of school systems for well-planned changes that may be related to the design of desirable school buildings to meet the goals of school education, school structure that is favourable for implementing the new school curriculum; school curriculum content that fits the needs of learners to be competitive in the digital world; and pedagogy that is appropriate to foster attributes of learners for meeting the needs in the digital world. The panel is expected to provide participants with insights related to preparing school learners living in the digital era to accommodate changes incurred in the emergence of digitalization in every aspect of daily life.

References


PANEL 3  *Multi-sensorial and Cross-disciplinary Innovative Design of Game-based Learning*

**ABSTRACT**  The intention of the panel, *Multi-sensorial and Cross-disciplinary Innovative Design of Game-based Learning*, is to showcase a wide array of game-based learning with their essential instructional concerns and ultimate learning goals of the design from the aspects of instructional design, gaming system development, learning objectives, and research analysis of various types and models of game-based learning. Panelists with rich teaching and research experiences in Hong Kong, Singapore, Taiwan, and Canada introduce projects such as, ICER Moodle Plug-In, MEGA World, GASIL, Green City Blues, and STEAM Port, that explore the possibilities of GBL and discuss the educational implications.

**FACILITATOR**  Ju-Ling Shih, National University of Tainan, Taiwan

**PANELISTS**

Maiga Chang, Athabasca University, Canada  
Morris Jong, The Chinese University of Hong Kong, Hong Kong  
Ming-Fong Jian, National Central University, Taiwan
Multi-Sensorial and Cross-Disciplinary Innovative Design of Game-based Learning

Ju-Ling SHIH\textsuperscript{a}, Maiga CHANG\textsuperscript{b}, Siu-Yung Morris JONG\textsuperscript{c} & Mingfong JAN\textsuperscript{d}  
\textsuperscript{a}Department of Information and Learning Technology, National University of Tainan, Taiwan \textsuperscript{b}School of Computing and Information Systems, Athabasca University, Canada \textsuperscript{c}Department of Curriculum and Instruction, The Chinese University of Hong Kong, Hong Kong \textsuperscript{d}Graduate Institute of Learning and Instruction, National Central University, Taiwan  
\textsuperscript{*} juling@mail.nutn.edu.tw

Abstract: The intention of the panel, Multi-sensorial and Cross-disciplinary Innovative Design of Game-based Learning, is to showcase a wide array of game-based learning with their essential instructional concerns and ultimate learning goals of the design from the aspects of instructional design, gaming system development, learning objectives, and research analysis of various types and models of game-based learning. Panelists with rich teaching and research experiences in Hong Kong, Singapore, Taiwan, and Canada introduce projects such as, ICER Moodle Plug-In, MEGA World, GASIL, Green City Blues, and STEAM Port, that explore the possibilities of GBL and discuss the educational implications.

Keywords: Game-based Learning, Gamification, Cross-discipline, Multi-sensorial, Instructional Design

1. Introduction

The goal of this panel is to create an interactive game-based learning discussion focusing on the instructional design, gaming system development, learning objectives, and research analysis of various types and models of game-based learning. The panel starts with a short interaction with the audience using mini-technologies surveying audience’s background and GBL practice experiences, as well as the issues they concern the most. In the following stage, the panelists showcase the innovative designs of game-based learning, and discuss related issues with the audience.

2. Innovative Designs of Game-based Learning

2.1 ICER Moodle Plug-In and MEGA World  
(Maiga CHANG)

Wu and Elliott (2008) point out that students might not appreciate if educational rewards are not useful for them as Kohn (1999) has also found that symbolic educational rewards may not meaningful for students. However, most of educationists and researchers also agree that the use of cash or money as rewards is not really a good choice and might need to avoid (McNinch, 1997). Although studies show different results of the effectiveness of rewards, rewards do help to get students motivated in learning and improve their learning performance for at least a short term (Winefield, Barnett, & Tiggemann, 1984; Witzel & Mercer, 2003). Card games are common to be seen and welcome by students in different ages and even for adults. But most of card games are commercial ones and difficult to be used by teachers as educational rewards due to teachers cannot access all kinds of cards for free.

The research team has developed a web-based discipline independent trading card game for teachers to use in any of their courses (Chen, Kuo, Chang, & Heh, 2009). Trading card game (TCG) is also called Collectible Card Game or Customizable Card Game, has hundreds of cards which is more than traditional card game such as poker’s 56 cards and such characteristic is extremely important for
teachers to use in-game cards as rewards – as students may lose their interest once they have seen and collected all cards. The web-based game is an open access trading card game which allows teachers to give any in-game card to students according to their performances in different learning activities (e.g., classroom participation, discussions, assignments, quiz, exams, etc.).

The research team has conducted a six-week experiment with a hundred and seventy-two fifth-grade students. In the experiment the students are divided into two groups and both of groups are asked to use a web-based English vocabulary system. The only difference is the students in experiment group are introduced the trading card game and can receive in-game cards as rewards according to their performance at the end of using the web-based learning system every time. The results showed that the use of the in-game cards as educational rewards not only engages students to use the vocabulary learning system but also improves their learning outcome (Chen, Kuo, Chang, & Heh, 2017).

The ICER (In-game Card as Educational Reward) Moodle plug-in aims to evaluate students’ performance on learning activities in Moodle and delivers them in-game items as rewards (Chen et al., 2017). Whenever students complete learning activities (e.g., assignments, quizzes, and discussions), the reward plug-in decides whether or not an educational reward which they can use in the game should be given according to the predefined criteria their teacher set. When students have better performance in terms of doing learning activities, they will receive more powerful in-game items from the reward plug-in. With these powerful in-game items’ help, students can have more fun in the gameplay or even show-off the items that they have to other students (Chen, Chang, Kuo, & Heh, 2016). For this reason, students may put more efforts on doing their homework and may be actively participated in the discussions in the class for getting better rewards.

MEGA World (Multiplayer Educational Game for All) is a web-based multiplayer educational game platform which supports any languages and is capable of access any existing external resources (e.g., multimedia, materials, online meetings, etc.)(Chang & Kinshuk, 2010). Teachers can create their virtual worlds as well as create learning and assessment activities (i.e., quests in the game) for students. Students can learn specific knowledge and reach the learning goal by taking and solving those quests while playing (Kuo, Chang, Kinshuk, & Liu, 2010). MEGA World currently has seven types of quests include greeting, item collection and delivery, sorting, treasure hunting and digging, calculation, fill-in-the-blank, and short answer quest type. Teachers are also provided editors like map and quest editors so they can freely expand their game worlds and design quests in different levels for different subjects/topics according to their own teaching plan.

Both of the ICER Moodle plug-in and MEGA World are discipline independent and can be adopted by teachers who are teaching different courses. Furthermore, the virtual worlds created by different teachers can be bridged and students can further learn, travel and play in the different worlds through portals arranged by their teachers. When more virtual worlds created and designed by the teachers in a school, students would be engaged to perform better while solving quests since they might want to be seen on the ranking lists, in MEGA World, such like Most Richest Players and Highest Level Players. On the other hand when more courses adopt and enable the ICER Moodle plug-in, students would be engaged to participate learning activities, in the real world, such like homework and classroom discussions since they might want to have more powerful cards to win the match or have their card collection book complete. Therefore, if multiple courses in a school could adopt either ICER, MEGA World, or both, then the effect of the game-based learning and rewarding mechanism might be kept stimulated student’s learning motivation and could be carried to the followed courses.

2.2 Gamifying Authentic Social Inquiry Learning (Siu-Yung Morris JONG)

Social inquiry learning is considerably adopted in social and humanities education. It emphasizes on students’ inquiry into humans and their relationships with the “societal” world from multiple perspectives, values and interests (Stripling, 2008). Thus, engaging students to interact with real-life, real-world contexts in a first-hand manner during the course of social inquiry learning is important. This pedagogic paradigm is termed “Authentic Social Inquiry Learning” (ASIL).
Curiosity is the best driving force for learning; keeping learners curious via engaging them in game-based activities is a desirable approach to education (Papert, 1993; Piaget, 1970). Gamification refers to the use of game elements in non-gaming contexts. Dominguez et al. (2013) further define gamification as the incorporation of game mechanics into non-gaming software applications to promote user experience and engagement.

Guided by Stripling’s (2008) social inquiry model and Dominguez’s (2013) gamification model, we have developed a GPS-based gamified learning mobile application for engaging and supporting students to pursue context-aware ASIL in outdoor environments. During the process of inquiry, based on the GPS-signal received, context-aware scaffolds for guiding students to accomplish the inquiry tasks will pop up according to their physical position in the learning site.

A study was carried out to investigate the learning effectiveness of the proposed “gamified” ASIL (GASIL) approach. We found that GASIL had different degrees of significant positive effects on different academic-achieving students. The findings provide grounds for a wider adoption of GASIL in school education.

2.3 How Do Pre-Service Teachers Design Learning Activities for Collaborative Argumentation with a Role-playing Game (Mingfong JAN)

In his seminal work Situated Language and Learning: A critique of traditional schooling, James Paul Gee (2004) coins the term “content fetish” to ridicule the 20th century mainstream education systems that foreground the importance of contents while ignoring skills that are critical for the 21st century globalized society, such as collaboration, communication, critical thinking, and creativity. In recent years, game-based learning was promoted as an approach that will innovate learning, and even solve 21st century learning challenges (Shaffer, Squire, Halverson, & Gee, 2005). Game researchers have articulated why and how games can engage players in higher-order thinking skills (e.g., Gee, 2003; Squire, 2006; Shaffer, 2006), conducted ethnographic research on games to understand how players think and collaborate (e.g., Lisk, Kaplancali, & Riggio, 2012; Steinkuehler, 2004) or led research initiatives that innovate learning with games (e.g., Kafai, 2006; Klopfer & Squire, 2008; Squire & Jan, 2007). According to such arguments and studies, games will undoubtedly play a critical role curing the content fetish of the 20th century education system.

Challenges arise when games that foster 21st century competencies are brought to schools (Jan, Tan, & Tan, 2015). In schools, teachers become a major director and user of the game system while this is not the case in out-of-school games. It is often teachers, instead of students (players), who decide what games will be played and how games should be played, therefore leading to the transformation of play. Consequently, how teachers perceive the affordances of games and how they design learning activities around games emerge as critical questions for game-based learning conducted in schools. Such questions are particularly relevant for games designed to foster 21st century skills, such as problem-solving and collaborative argumentation, because they don’t resemble textbooks and other content-delivery technologies.

Another issue that might stagnate game-based learning is pertinent to teacher education - how teachers are taught to become teachers, and how they are taught to use instructional materials for learning. Teacher education in Taiwan predominantly engages teachers as content experts, expecting teachers to teach domain-specific contents with both traditional and emerging pedagogies and technologies. The official teacher education programs have yet responded to the needs for teaching 21st century competencies. Due to the above challenges, how will teachers use games to foster 21st century skills? We begin with the inquiry with pre-service teachers.

To investigate the above question, we engage six pre-service teachers to design learning activities with Green City Blues - a role-playing card game designed for collaborative argumentation. With a qualitative case study approach (Stake, 1995), we interrogate how they design learning activities with Green City Blues and how they perceive the affordances of the game. We collect lesson plans designed by the six teachers, code the data with an open-coding technique in order to identify emerging themes regarding teachers’ design thinking.
Findings suggest the following themes. In terms of pedagogy, participants propose to teach 21st century competencies based on how they were taught in schools – content mastery and direct instruction. Viewing learning as acquisition, they look for contents to be taught from Green City. Although some foreground the role of being facilitators, they could not articulate how and what to facilitate. In terms of the perceived affordances of the game, they don’t view the game as providing an authentic context for making arguments collaboratively. Instead, they think the game was difficult to use because there were no textbook-like canonical knowledge. The findings suggest a critical issue with pedagogy and epistemology when it comes to using 21st century games—the pre-service teachers employ an acquisition epistemology to teach 21st century skills, a misalignment between the designed affordances of Green City Blues and proposed pedagogy from pre-service teachers.

The 20th century education system is a fortified system that is resilient to changes, even under the pressure of global competition. In 20th century education system, teacher education, curricula, subject contents, configuration of classrooms, instructional technologies, pedagogies, and assessments all have been designed and calibrated to ensure the maximum efficiency of content delivery. Through the enduring calibration and practices, a “textbook-learning culture” is reproduced and reinforced to become a dominant lens for not only schooling, but also views on learning and teaching. Teacher education, likewise, is conceptualized based on teaching the mastery of disciplinary content (Gee, 2004). Through this study, we recognize that pre-service teachers’ epistemology in teaching might be one of the major obstacles for successful implementation of game-based learning in schools.

2.4 STEAMing the Ships for the Great Voyage: <STEAM Port>
(Ju-Ling SHIH)

This project started with an innovative complex board game based on a 3-dimensional physical pop-up map with the integration of technologies and media, such as near field communication (NFC), 2D mobile games, and augmented reality. Now, it is developed into an interdisciplinary learning model that puts students in a cultural historical context with a classroom-size world map, in which they would apply their knowledge of history, geography, math, physics, mechanics and natural sciences to play a magnified board game.

In STEAM Port, a maker game based on the history of the Age of Discovery, players represent different countries that sail to Asia for spices and other goods. In the game, the students first transform robots into ships in the maker class, and then use programming skills to set up the parameters of the ship according to the ship powers of respective countries. Then the players use their knowledge about the history, acquire the characteristic spices for each location, get acquaintance to the trading conditions, and compete in groups to retrieve designated spices by controlling robots with computers. With every successful returning freight, players win chances to sail further or upgrade their ships; or else, they use smart phones to remote control their ships for battles. The game involved interdisciplinary learning in STEAM: Science (sailing and finding channels), Technology (programming), Engineering (wiring robots), Art (crafting for ships), and Mechanics (building robotic ships).

In the learning scenario, students’ critical thinking, creative thinking, computational thinking, and problem-solving abilities, as well as cooperative and competitive gaming strategies would be used and enhanced. These are the 21st century competencies students should have that encourage students to transform crystal learning contents into fluent competencies so they would be apply them in their future lives.

It is designed to be sustainable, extensible, and feasible for formal education to adopt for classroom interactions or extra-curricular activities. Students can be immersed in the cultural historical context which is produced by the amalgamation of real and virtual spheres. In-between the world map, ship crafts, stories, and games, students experience the joy of traveling in spaces and times. This project also provides teachers a model for flipped classroom in primary and secondary schools with exemplarily teaching plans to guide the uses of complex board game.
3. Discussion

Issues of game-based learning would be discussed with the panelists and audience including,

a. What are the instructional concerns of the mentioned GBL design? Do they or do they not work when they are in practices, and why? What should teacher pay attention to?

b. What are the learning outcomes of the mentioned GBL design? Anything special or interesting that you find from your instructional practices?

c. What are the research and analysis designs of the mentioned GBL practices? Are they or are they not able to see students’ changes in various ways?

d. In what way do you think can make the GBL design accessible, feasible, and sustainable?

4. Conclusion

The intention of the panel, Multi-sensorial and Cross-disciplinary Innovative Design of Game-based Learning, is to showcase a wide array of game-based learning with their essential instructional concerns and ultimate learning goals of the design. In a way, "multi-sensorial" can be widely defined as using eyes(detections), hands(maker), olfactions, body movements (kinect/mobile), etc. that can include all kinds of virtual, real, blended, technology-enhanced interactive games/learning; and "cross-disciplinary" can be anything that come across to any subject in whichever education level. This panel has invited panelists with teaching and research experiences in Hong Kong, Singapore, Taiwan, and Canada. A lively and interactive talk among panelists and the audience are brought about so the issues are discussed in depth.

Acknowledgements

Thanks to all members in the laboratories under the guidance of all the panelists listed above.

References


# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Paper Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C1: Artificial Intelligence in Education/Intelligent Tutoring System (AIED/ITS) and Adaptive Learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigating the effects of Cognitive and Metacognitive Scaffolding on Learners using a Learning by Teaching Environment</td>
<td>Cristina Dumumaya, Michelle Banawan, Ma. Mercedes T. Rodrigo, Amy Ogan, Evelyn Yarzebinski and Noburo Matsuda</td>
<td>1</td>
</tr>
<tr>
<td>Learning arithmetic word problem structure with a picture combination application in Kindergarten</td>
<td>Pedro Gabriel Fonteles Furtado, Tsukasa Hirashima, Yusuke Hayashi and Kazumasa Maeda</td>
<td>11</td>
</tr>
<tr>
<td>Framework for Building a Thinking Processes Analysis Support System: A Case Study of Belief Conflict Thinking Processes</td>
<td>Yuki Hayashi, Kazuhisa Seta and Mitsuru Ikeda</td>
<td>21</td>
</tr>
<tr>
<td>Usability and Learning Effect Evaluations of an Electrical Note-Taking Support System with Speech Processing Technologies</td>
<td>Hiromitsu Nishizaki and Yusuke Narita</td>
<td>31</td>
</tr>
<tr>
<td>A Student Placement Predictor for Programming Class Using Class Attitude, Psychological Scale, and Code Metrics.</td>
<td>Ryosuke Ishizue, Kazunori Sakamoto, Hironori Washizaki and Yoshiaki Fukazawa</td>
<td>40</td>
</tr>
<tr>
<td><strong>Short papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding Support System for Causal Relationship in Historical Learning</td>
<td>Tomoko Kojiri, Fumito Nate and Keitaro Tokutake</td>
<td>50</td>
</tr>
<tr>
<td>Extraction of Relationships between Learners’ Physiological Information and Learners’ Mental States by Machine Learning</td>
<td>Yoshimasa Tawatsuji, Tatsuro Uno, Keita Okazaki, Siyuan Fang and Tatsunori Matsui</td>
<td>56</td>
</tr>
<tr>
<td>Predicting Student Carefulness within an Educational Game for Physics using Support Vector Machines</td>
<td>Michelle Banawan, Ma. Mercedes Rodrigo and Juan Miguel L. Andres</td>
<td>62</td>
</tr>
<tr>
<td>GPT: A Tutor for Geometry Proving</td>
<td>Roscoe Nealle Alibusan, Laurence Nicholas Foz, Paolo Vittorio Merle, Ethel Chua Joy Ong and Minie Rose Lapinid</td>
<td>68</td>
</tr>
<tr>
<td>Inquiry-based Support System to Improve Intention Sharing Skills</td>
<td>Natsumi Mori, Yuki Hayashi and Kazuhisa Seta</td>
<td>74</td>
</tr>
<tr>
<td>A Tool for data acquisition of thinking processes through writing</td>
<td>Wasan Na Chai, Taneth Ruangrajitpakorn and Thepchai Supnithi</td>
<td>80</td>
</tr>
<tr>
<td>An Artificial Intelligence Approach to Identifying Skill Relationship</td>
<td>Tak-Lam Wong, Yuen Tak Yu, Chung Keung Poon, Haoran Xie, Fu Lee Wang and Chung Man Tang</td>
<td>86</td>
</tr>
</tbody>
</table>
Posters

Development of a System to Construct Explanation for Physics Phenomenon through Thought Experiment
Miki Matsumuro, Kazuhisa Miwa and Yuiko Tombe
Towards a Virtual Peer to Write Stories with Children
Hans Gabriel Chua, Geraldine Elaine Cu, Chester Paul Ibarrientos, Moira Denise Paguiligan and Ethel Ong
Prospects in Modeling Reader's Affect based on EEG Signals
Kristine Kalaw, Ethel Ong and Judith Azcarraga
Proposal of an Intelligent Tutoring System for procedural learning with Context-aware Dialogue
Jose Paladines and Jaime Ramirez
Improvement of the Situational Dialog Function and Development of Learning Materials for a Japanese Dictogloss Environment
Satoru Kogure, Kaito Okugawa, Yasuhiro Noguchi, Tatsuhiro Konishi, Makoto Kondo and Yukihiro Itoh
Semantically Enhanced Gaze-aware Historical Cartoons to Encourage Historical Interpretation
Daiki Muroya, Yuki Hayashi and Kazuhisa Seta
Developing a Geometric Proof Problem-Solving Support System Utilizing Card Selection
Megumi Kurayama
Analyzing Novice Programmers’ EEG Signals using Unsupervised Algorithms
Vanlalhruaii Swansi, Tita Herradura and Merlin Suarez
Evaluation of mathematics knowledge level through personalized learning exercise based on the Adaptive Tests
Yan Liu, Chuxin Fu and Xiaoqing Gu

C2: Computer-supported Collaborative Learning (CSCL) and Learning Sciences

Full Papers

Discovering Dynamics of an Idea Pipeline: Understanding Idea Development within a Knowledge Building Discourse
Alwyn Vwen Yen Lee and Seng Chee Tan
Using CSCL to Conceptualize Disability Toward Inclusive Education Design
Christopher P. Ostrowski
Measuring Process and Outcome of the Scientific Argumentation in a CSCL Environment
Wenli Chen and Chee Kit Looi
Case-based Portraits of Contrasting Micro-Interaction Processes During Online Assessment of Collaborative Problem Solving
Johanna Pöysä-Tarhonen, Esther Care, Nafisa Avwal and Päivi Hääkinen
DBCollab: Automated Feedback for Face-to-Face Group Database Design
Vanessa Echeverria, Roberto Martinez-Maldonado, Katherine Chiluiza and Simon Buckingham Shum
Examining Student Learning of Engineering Estimation from METTLE
Aditi Kothiyal and Sahana Murthy
Social Media Facilitated Group Performance: An Investigation of Tie Strength in Grouping
Cong Qi

Short papers

Impact of Both Prior Knowledge and Acquaintanceship on Collaboration and Performance: A Pair Program Tracing and Debugging Eye-Tracking Experiment
Maureen Villamor and Ma. Mercedes Rodrigo
A Toolkit for Action: Translating Theory into Practice
Lucila Carvalho and Pippa Yeoman

Exploratory Analysis of Discourses between Students Engaged in a Debugging Task
Ma. Mercedes T. Rodrigo

Analyzing a Practical Implementation of Training Metacognition through Solving Mathematical Word Problems
Tama Duangnamol, Thepchai Supnithi, Gun Srijuntongsiri and Mitsuru Ikeda

Designing the EMBeRS Summer School: Connecting Stakeholders in Learning, Teaching and Research
Kate Thompson, Antje Danielson, David Gosselin, Simon Knight, Roberto Martinez-Maldonado, Roderic Parnell, Deana Pennington, Julia Svoboda-Gouvea, Shirley Vincent and Penny Wheeler

Why Learners Fail in MOOCs? Investigating the Interplay of Online Academic Hardiness and Learning Engagement among MOOCs Learners
Tonny Meng-Lun Kuo, Chin-Chung Tsai and Jyun-Cheng Wang

Toddlers Testing DDMM: Evaluation Results and Ideas towards Creating Better Learning Environments for Small Children
Christine Steinmeier and Dominic Becking

Effects of Peer Interaction on Web-Based Computer Programming Learning
Nuttaphat Arunoprayoch, Chih-Hung Lai, Pham-Duc Tho and Jing-San Liang

Posters

Exploring Lag Times in a Pair Tracing and Debugging Eye-Tracking Experiment
Maureen Villamor and Ma. Mercedes Rodrigo

The Research of Interaction Performance of Intercultural Communication in Computer-supported Collaborative Learning
Yi Wei, Li-Jie Wu and Yi-Ling Hu

Collaborative Learning in Elementary Science Supported by Learning by Inquiry and Augmented Reality
Sie Wai Chew, I-Hsiu Lin, Kinshuk and Nian-Shing Chen

Collaborative Inductive Problem Solving Using an ICT Tool in an Elementary Science Classroom
Hirokazu Kawano and Shu Matsuura

The Design of a Portfolio-Based Reading Conversation Platform
Tzu-Chao Chien, Zhi-Hong Chen and Tak-Wai Chan

Improving Reading Comprehension using the Cooperative Mind Mapping Summary Strategy
Gloria Yi-Ming Kao, Ju-I Cheng and Chin-Chung Tsai

C3: Advanced Learning Technologies (ALT), Learning Analytics and Digital Infrastructure

Full Papers

Resource Description Framework (RDF) Models for Representing the Revision Process in Research Support Systems
Harriet Ocharo and Shinobu Hasegawa

Extracting Implicit Suggestions from Students’ Comments – A Text Analytics Approach
Venky Shankararaman, Swapna Gottipati and Jeff Lin Rongsheng

Semi-Discovery Learning Support System for Analogical Reasoning in High-School Physics
Yasuhiro Noguchi, Takeshi Enokida, Tatsuhiro Konishi and Yukihiro Itoh
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using Data Analytics for Discovering Library Resource Insights – Case from Singapore Management University</td>
<td>Lu Ning, Song Rui, Dina Heng Lie Gwek, Swapna Gottipati and Aaron TAY</td>
<td>280</td>
</tr>
<tr>
<td>A Bibliometric Analysis of 15 Years of Research on Open Educational Resources</td>
<td>Xiaochen Wang, Mengrong Liu, Qianhui Li and Yuan Gao</td>
<td>296</td>
</tr>
<tr>
<td>Design and Implementation of a Pedagogic Intervention Using Writing Analytics</td>
<td>Antonette Shibani, Simon Knight, Simon Buckingham Shum and Philippa Ryan</td>
<td>306</td>
</tr>
<tr>
<td>Multimodal Interaction Aware Platform for Collaborative Learning</td>
<td>Aoi Sugimoto, Yuki Hayashi and Kazuhisa Seta</td>
<td>316</td>
</tr>
<tr>
<td>Using Network-Text Analysis to Characterise Learner Engagement in Active Video Watching</td>
<td>Tobias Hecking, Vania Dimitrova, Antonija Mitrovic and H. Ulrich Hoppe</td>
<td>326</td>
</tr>
<tr>
<td>Synergizing Online Group Knowledge</td>
<td>Yen-An Shih and Ben Chang</td>
<td>336</td>
</tr>
<tr>
<td>A Presentation Avatar for Self-Review</td>
<td>Keisuke Inazawa and Akihiro Kashihara</td>
<td>345</td>
</tr>
<tr>
<td>Cognitive Investigation of Dynamic Educational Presentation toward Better Utilization of Presentation Characteristics</td>
<td>Yasuhisa Okazaki and Atsushi Yoshikawa</td>
<td>355</td>
</tr>
<tr>
<td>The Journey to Improve Teaching Computer Graphics: A Systematic Review</td>
<td>Thomas Suselo, Burkhard Wuesche and Andrew Luxton-Reilly</td>
<td>361</td>
</tr>
<tr>
<td>Designing a &quot;Three Rings&quot; Theory Framework for Electronic Schoolbag</td>
<td>Baoyuan Yin, Fati Wu, Shihua Huang and Shan Jia</td>
<td>367</td>
</tr>
<tr>
<td>Interaction between Standardisation and Research in Drafting an International Specification on Learning Analytics</td>
<td>Tore Hoel and Wei Qin Chen</td>
<td>373</td>
</tr>
<tr>
<td>An Evaluation of Elementary Students’ Ability of Problem Solving in Information Processing</td>
<td>Lishan Zhang, Jing Wang, Zijun Zhuang and Baoping Li</td>
<td>379</td>
</tr>
<tr>
<td>Algorithm Learning by Comparing Visualized Behavior of Programs</td>
<td>Daiki Ihara, Satoru Kogure, Yasuhiro Noguchi, Koichi Yamashita, Tatsuhiro Konishi and Yukihiro Itoh</td>
<td>385</td>
</tr>
<tr>
<td>Analyzing the E-learning Video Environment Requirements of Generation Z Students using Echo360 Platform</td>
<td>Swapna Gottipati and Venky Shankararaman</td>
<td>391</td>
</tr>
<tr>
<td>Train-For-Life: On-Line Interactive Training for Industry Learners</td>
<td>Bashar Barmada and Nilufar Baghaei</td>
<td>397</td>
</tr>
<tr>
<td>Detect Students’ Academic Emotions in Classroom: Measurement, Self-perception and Manifested Behaviors</td>
<td>Yan Liu, Menghua Hu and Xiaoping Gu</td>
<td>403</td>
</tr>
<tr>
<td>Observing the Degree of Distortion in Coordinated Motor Actions</td>
<td>Takehiko Yoshikawa, Kohta Sugawara, Kenji Matsuura, Stephen Karungaru and Naka Gotoda</td>
<td>415</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Learning Environment for Recursive Functions by Visualization of Execution Process</td>
<td>Raiya Yamamoto, Yasuhiro Anzai, Satoru Kogure, Yasuhiro Noguchi, Koichi Yamashita, Tatsuhiko Konishi and Yukihiro Itoh</td>
<td>421</td>
</tr>
<tr>
<td>Can Distributed Practice Improve Students’ Efficacy in Learning their First Programming Language?</td>
<td>Qiujie Zhang, Lishan Zhang, Baoping Li, Ling Chen, I-Han Hsiao and Fati Wu</td>
<td>427</td>
</tr>
<tr>
<td>Cross Analytics of Student and Course Activities from e-Book Operation Logs</td>
<td>Atsushi Shimada and Shinichi Konomi</td>
<td>433</td>
</tr>
<tr>
<td>Students’ Performance Prediction Using Data of Multiple Courses by Recurrent Neural Network</td>
<td>Fumiya Okubo, Takayoshi Yamashita, Atsushi Shimada and Shinichi Konomi</td>
<td>439</td>
</tr>
<tr>
<td>A Case Study of Interactive Learning Environment for Building Structure of Arithmetic Word Problem in Language Delay</td>
<td>Sho Yamamoto and Tsukasa Hirashima</td>
<td>445</td>
</tr>
<tr>
<td>A Learning Support System for Integrated Motor Skill by Organized Training Stages</td>
<td>Kohta Sugawara, Takehiko Yoshikawa, Kenji Matsuura, Stephen Karungaru and Naka Gotoda</td>
<td>451</td>
</tr>
<tr>
<td>Enhancing Students’ Critical Reading Fluency, Engagement and Self-Efficacy using Self-Referenced Learning Analytics Dashboard Visualizations</td>
<td>Christin Jonathan, Jennifer Pei-Ling Tan, Elizabeth Koh, Imelda Caleon and Siu Hua Tay</td>
<td>457</td>
</tr>
<tr>
<td>Coloring Strategy Combined with Three-Dimensional Animation: Does the Mental Strategy Fits Everyone?</td>
<td>Jiayu Chu, Yiling Hu, Shiyu Wang and Xiaoqing Gu</td>
<td>463</td>
</tr>
<tr>
<td>Development of Alternative Conception Diagnostic System based on Item Response Theory in MOOCs</td>
<td>Yu-Cheng Cheng, Jian-Wei Tseng, Nen-Fu Huang, Chia-An Lee and Meng-Lun Kuo</td>
<td>469</td>
</tr>
<tr>
<td>Capturing Changes and Variations from Teachers’ Time Series Usage Data</td>
<td>Longwei Zheng, Rui Shi and Xiaoqing Gu</td>
<td>475</td>
</tr>
<tr>
<td>TGlass: A Custom-made Wearable Promoting Accessibility for Tetraplegic</td>
<td>Maikon Soares, Felipe Bede, Thiago Araujo, Daniel Cavalcante, Adriano Freitas, Zulina Carvalho and Francisco Oliveira</td>
<td>481</td>
</tr>
</tbody>
</table>

**Posters**

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Reality based Learning Support System for Mental Rotation</td>
<td>Midori Nakano, Yukihiro Matsubara, Masaru Okamoto and Noriyuki Iwane</td>
<td>487</td>
</tr>
<tr>
<td>Monitoring System to Help an e-Learning Institution to Manage Tutors and Student Data Retrieved from Moodle</td>
<td>Elza Monier, Fátima Gatinho, Ana Emilia Oliveira, Aldrea Rabelo and Rômulo Martins</td>
<td>490</td>
</tr>
<tr>
<td>A Study on Prediction of Academic Performance based on Current Learning Records of a Language Class using Blended Learning</td>
<td>Byron Sanchez, Xiumin Zhao, Takashi Mitsuishi and Terumasa Aoki</td>
<td>493</td>
</tr>
<tr>
<td>The Development of a Simulation to Support Authentic Observation in Precipitation Reaction</td>
<td>Chang Youn Lee and Hun-Gi Hong</td>
<td>496</td>
</tr>
<tr>
<td>Kanji Learning Support with Feedback based on Haptic and Pseudo-Haptic</td>
<td>Takanori Kono, Yukihiro Matsubara and Masaru Okamoto</td>
<td>499</td>
</tr>
<tr>
<td>Discourse Analysis of Teachers’ Commentary on Students</td>
<td>Rui Shi, Longwei Zheng and Xiaoqing Gu</td>
<td>502</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Discovering Teachers’ In-Class ICT Usage With Frequent Closed Sequence Mining</td>
<td>Bingcong Wu, Longwei Zheng and Xiaoqing Gu</td>
<td>505</td>
</tr>
<tr>
<td>Support for the Cycle of Task Extraction, Goal Setting and Assessment in Research Activities</td>
<td>Shigeki Ohira, Naoya Morita and Katashi Nagao</td>
<td>508</td>
</tr>
<tr>
<td>AR-based Inorganic Chemistry Learning Support System using Mobile HMD</td>
<td>Masaru Okamoto, Tsukasa Ishimura and Yukihiro Matsubara</td>
<td>511</td>
</tr>
<tr>
<td>Analysis on Students’ Usage of Highlighters on E-textbooks in Classroom</td>
<td>Yuta Taniguchi, Fumiya Okubo, Atsushi Shimada and Shinichi Konomi</td>
<td>514</td>
</tr>
<tr>
<td>Taxonomy for Teacher-Actionable Insights in Learning Analytics</td>
<td>Elizabeth Koh and Jennifer Pei-Ling Tan</td>
<td>517</td>
</tr>
<tr>
<td><strong>C4: Classroom, Ubiquitous, and Mobile Technologies Enhanced Learning (CUMTEL)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Architecture and Predictive Experiment for an Automatic Learning Support Function on Classroom Response Systems</td>
<td>Kozo Mizutani</td>
<td>520</td>
</tr>
<tr>
<td>Enhancing Seamless Learning Using Learning Log System</td>
<td>Noriko Uosaki, Hiroaki Ogata and Kousuke Mouri</td>
<td>529</td>
</tr>
<tr>
<td>Students’ Participative Stances and Knowledge Construction in Small Group Collaborative Learning with Mobile Instant Messaging Facilitation</td>
<td>Ying Tang, Khe Foon Hew and Gaowei Chen</td>
<td>539</td>
</tr>
<tr>
<td>Learning Behavioral Pattern Analysis based on Students’ Logs in Reading Digital Books</td>
<td>Chengjiu Yin, Noriko Uosaki, Hui-Chun Chu, Gwo-Jen Hwang, Jau-Jian Hwang, Itsuo Hatono, Etsuko Kumamoto and Yoshiyuki Tabata</td>
<td>549</td>
</tr>
<tr>
<td>A Case study of Evaluation of Learners’ Acceptance of AR_H2O2 System</td>
<td>Tao Wang, Shan Jia, Jinrui Dai, Manli Lu, Su Cai and Feng-Kuang Chiang</td>
<td>558</td>
</tr>
<tr>
<td>An Empirical Case on Integration of Immersive Virtual Environment into Primary School Science Class</td>
<td>Jian Sun, Hao Li, Zhanhao Liu, Su Cai and Xiaowen Li</td>
<td>566</td>
</tr>
<tr>
<td>Online Synchronous Discussion in Face-to-Face Classroom Based on WeChat</td>
<td>Manli Lu, Guang Chen and Shuxian Ouyang</td>
<td>576</td>
</tr>
<tr>
<td>Teachers’ Concerns about Adopting Interactive Spherical Video-based Virtual Reality</td>
<td>Jie Geng, Tsun Hin Eric Luk and Siu Yang Morris Jong</td>
<td>586</td>
</tr>
<tr>
<td>How Designer Think About Designing an Augmented Reality App for the Study of Central Nervous System</td>
<td>Fadzil Saleh Mohamad Rofie and Mas Nida Md. Khambhari</td>
<td>594</td>
</tr>
<tr>
<td>A Comparison of Different Types of Learning Activities in a Mobile Python Tutor</td>
<td>Geela Venise Firmalo Fabric, Antonija Mitrovic and Kourosh Neshatian</td>
<td>604</td>
</tr>
<tr>
<td><strong>Short papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Case study of Young Children’s Use of iPad for Digital Storytelling for a Study of Self</td>
<td>Leigh Disney and Gretchen Geng</td>
<td>614</td>
</tr>
<tr>
<td>Online Responses towards the Impact of Hand Held Devices on Children’s Social and Emotional Development</td>
<td>Gretchen Geng and Leigh Disney</td>
<td>620</td>
</tr>
</tbody>
</table>
Applying Pedagogy to the Design of Software for Helping Students Learn Equation Solving

Daphne Robson 626

The Development of Mobile Learning CPD Modules to Improve the Management of Respiratory Diseases

Georgina Orsborn, Teresa Demetriou, Kerri Arcus and Elizabeth Asbury 632

The Effectiveness of Media Platforms on Reading Comprehension: A Meta-analysis

Bing Xu, Guang Chen, Yuting Sun and Ronghuaui Huang 638

Explore the Impact of Collaborative Tendencies in the Flipped Classroom on Taking Basketball Teaching

Tosti H. C. Chiang 644

Teaching Influence for Perceived Usefulness of Interactive Whiteboard - Based on the Perspective of College Students

Peng Zhou, Zhexu Liu and Jian Xu 650

Designing Boundary Activity for Mobile Learning in Science Inquiry

Daner Sun and Chee-Kit Looi 656

Responsive eBook based on the Principles of Educational Interfaces

Ana Emilia Oliveira, Elza Monier, Dilson Rabelo, Fátima Gatinho, Aldrea Rabelo and Marcelo Henrique Monier Alves Júnior 662

Encouraging System for Teaching Assistants to Advise Students during Programming Exercises

Yuuki Yokoyama and Hironori Egi 665

Improving Primary Students’ Problem Solving Skills in Science Learning in a Seamless Learning Environment

Yanjie Song and Ka Man Lung 671

Blockino: a Tool with an Emphasis on Educational Robotics Assisting the Teaching of Programming Logic

Higo Sampaio, Mauro Silva, Aldrea Rabelo, Carla Marques and Marcelo Monier 676

Effects of Prior Knowledge of High Achievers on Use of e-Book Highlights and Annotations

Misato Oi, Fumiya Okubo, Yuta Tanimuchi, Masanori Yamada and Shinichi Konomi 682

Promoting Extrinsic Motivation Based on Result of LMS Quiz

Yasutaka Asai and Hironori Egi 688

An Electronic ID System Using a Smart Phone

Jaewook Kim, Takashi Tachino and Yasuhiito Kishi 694

Social Network Analysis of Teacher’s Role in Students’ Online Discussion Community

Chih-Ming Chu 700

Exploring the Nature of Teacher’s Ongoing Feedback to Pupil using iPad

Weiyun Chen, Yaoqin Xue and Xiaojing Gu 706

Development of a Community-based Hazard Information Sharing System for Traditional Towns with Local Heritage

Shun Kozaki, Yasuhsa Okazaki, Hiroshi Wakaya, Yuku Hayashida, Byung-Won Min and Nobuo Mishima 712

Educational Effectiveness of a System for Scientific Observation of Animals in a Zoo

Yuji Tanaka, Ryohei Egusa, Yuuki Dobashi, Fusako Kusunoki, Etsuji Yamaguchi, Shigenori Inagaki and Tomoyuki Nogami 715

Leveraging an Existing Learning Management System for Alternative Learning

Mikaela Malit and Ma. Mercedes Rodrigo 718

Posters
Providing Regular Assessments and Earlier Feedback on Moodle in an Introductory Computer Science Course: A User Study  
Natalia Nehring, Simon Dacey and Nilufar Baghaei  
721

Study on Implementing Automated Classroom Performance System for Recording Student Attendance  
Daniel Mangalaraj and Shankar Subramanian  
724

Stream-based Reasoning for IoT Applications on Domain Education  
Marcelo Alves Júnior, Francisco Silva, Markus Endler, Vitor Almeida and Higo Sampaio  
727

A Meta Analysis: The Effectiveness of E-Schoolbag Use on Students’ Academic Achievement in China  
Menghua Hu and Yiling Hu  
730

Real-time Analysis of Digital Textbooks: What Keywords Make Lecture Difficult?  
Kousuke Mouri, Atsushi Shimada, Chengjiu Yin, Uosaki Noriko, Yachirawit Tengchaisri and Keiichi Kaneko  
733

<table>
<thead>
<tr>
<th>C5: Digital Game and Digital Toy Enhanced Learning and Society (GTEL&amp;S)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Papers</strong></td>
</tr>
</tbody>
</table>
| Designing an Intervention for Novice Programmers Based On Meaningful Gamification: An Expert Evaluation  
Jenilyn Agapito and Ma. Mercedes Rodrigo  
736 |
| Virtual Currency as Gamification for Learning in a Disaster Museum to Increase the Number of Revisitors  
Hiroyuki Mitsuhara and Masami Shishibori  
746 |
| “Go Kahoot!” Enriching Classroom Engagement, Motivation and Learning Experience with Games  
Sherlock Licorish, Jade Li George, Helen Owen and Ben Daniel  
755 |
| JobStar Online: Game-Based Learning on Smartphones to Promote Youth Career Education  
Toru Fujimoto, Yuki Fukuyama, Satoko Azami and Satoru Konno  
765 |
| Sufficiency Economy Philosophy-Based Mobile Game Application to Promoting Sustainability Understanding based on Inquiry Learning with Everyday Life Activities  
Charoenchai Wongwatkit, Muttakeen Che-Leah, Sasithorn Chookaew, Jintana Wongta, Chitphon Yachulawetkunakorn and Ratthakarn Na Phatthalung  
774 |
| **Short papers** |
| The Digital Interactive Learning Theater in the Classroom for Drama-based Learning  
Yu-Tzu Liu, Shang-Chiao Lin, Wei-Yi Wu and Gwo-Dong Chen  
784 |
| Learning Support System for Museum exhibits using Complex Body Movements –Enhancing Sense of Immersion in Paleontological Environment  
Mikihiro Tokuoka, Hiroshi Mizoguchi, Ryohei Egusa, Shigenori Inagaki, Fusako Kusunoki and Masanori Sugimoto  
790 |
| SATOYAMA: Simulating and Teaching Game Optimal for Young Children to Learn Vegetation Succession as Management of an Actual Forest  
Shuya Kawaguchi, Hiroshi Mizoguchi, Ryohei Egusa, Yoshiuki Takeda, Etsuji Yamaguchi, Shigenori Inagaki, Fusako Kusunoki, Hideo Funaoi and Masanori Sugimoto  
796 |
| Muse: A Musically Inspired Game To Teach Arrays and Linked Lists  
Vaishali Sharma, Raghib Musarrat, Sridhar Chimalakonda and Y Raghu Reddy  
802 |
| Healthy Kidney: An Educational Game for Health Awareness  
Sanda Hammedi, Fathi Essalmi and Maiga Chang  
808 |
Posters

The Effects of SDE Strategy-based Computer Games on Metacognitive Awareness  
Kai-Hsiang Yang, Chin-Yen Lu and Bou-Chan Lu  814

Scaffolding historical inquiry through a collaborative maker-based activity  
Benjamin Lille and Margarida Romero  817

A Study of Design Thinking Adaptation for Maker Education Process  
Peng Chen, Ding Li and Ronghuai Huang  820

A POE Strategy-Based Gaming Approach for Mathematics Learning  
Kai-Hsiang Yang, Hsiao-Hua Chen and Bou-Chuan Lu  823

Probing in-service Teachers’ Perceptions on TPACK-G and Acceptance of GBL  
Chung-Yuan Hsu, Jyh-Chong Liang, Ching Sing Chai and Chin-Chung Tsai  826

Game-based Narrative System for Student English Learning  
Zhi-Hong Chen  829

Exploring Teachers’ Pedagogical Design Thinking in Game-based Learning  
Mingfong Jan, Wan-Lin Yang and Michael Thomas  832

C6: Technology Enhanced Language Learning (TELL)

Full Papers

Can Conversational Agents Foster Learners’ Willingness To Communicate in a Second Language? : Effects of Communication Strategies and Affective Backchannels  
Emmanuel Ayedoun, Yuki Hayashi and Kazuhisa Seta  835

Multimodality in Language Education – Exploring the Boundaries of Digital Texts  
Anna-Lena Godhe and Petra Magnusson  845

Effectiveness of a Learning Design Combining Summary-speaking Self-study Using Mobile Application with Paired Reflection on Learners’ Speaking Process  
Kae Nakaya and Masao Murota  855

Discussion Course Model Using Online Educational Resources to Enhance EFL Learners’ Motivation and Critical Thinking  
Yuichi Ono and Ai Nakajima  865

Reading for Emotion with ICT Tools  
Ania Lian  874

Short papers

Automatic Classification of Teacher Feedback and Its Potential Applications for EFL Writing  
Gary Cheng, Shu-Mei Gloria Chwo, Julia Chen, Dennis Foung, Vincent Lam and Michael Tom  884

Automatic Question Generation System for English Exercise for Secondary Students  
Tasanawan Soonklang, Sunee Pongpinigpinyo, Weenawadee Muangon and Sirak Kaewjammong Sugimoto  890

Student Engagement with an Online Preenrolment English Course at a Japanese University  
Adam Smith and Andrew Johnson  896

Captioning Methods of Lecture Videos for Learning in English  
Veri Ferdiansyah and Seiichi Nakagawa  902
### Using Learning Analytics to Support Computer-Assisted Language Learning

Huiyong Li, Hiroaki Ogata, Tomoyuki Tsuchiya, Yubun Suzuki, Satoru Uchida, Hiroshi Ohashi and Shinichi Konomi

908

### An Approach to Accent Visualisation for the Reduction of Vowel Pronunciation Errors

Tom Anderson, Barry Reynolds and David Powers

914

---

### Posters

- **Chinese Grammatical Error Detection Using a CNN-LSTM Model**
  - Lung-Hao Lee, Bo-Lin Lin, Liang-Chih Yu and Yuen-Hsien Tseng
  - 919

- **Global Collaborative Learning Support System for Facilitator Collaboration: First Phase Development Report**
  - Yoshiko Goda, Masanori Yamada, Yumi Ishige and Junko Handa
  - 922

- **The Teleexistence Robot combing with Virtual Reality for Teaching English**
  - Wen-Chi Vivian Wu, Tosh Yamamoto, Cheng-Hao Hu and Rong-Jyue Wang
  - 925

- **Integration of Peer Assessment and Shadowing Strategies for Improving the Oral Performance of EFL Learners**
  - Siao-Cing Guo and Ting-Chia Hsu
  - 928

- **Effect Analysis of Students’ Learning Styles on Learning Experience with Lecture Videos Played at Different Playback Speeds**
  - Toru Nagahama and Yusuke Morita
  - 931

- **Word Error Rate as a Listenability Index for Learners of English as a Foreign Language**
  - Katsunori Kotani and Takehiko Yoshimi
  - 934

---

### C7: Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)

#### Full Papers

- **Science Teachers’ Engagement with ICT in Singapore: Different Perspectives**
  - Aik Ling Tan and Seng Chee Tan
  - 937

- **Participation and Psychological Ownership on Teachers’ Beliefs of a Cloud-Based VLE**
  - Joanne Sau-Ching Yim, Priscilla Moses and Alia Azalea
  - 946

- **A Learning Support Method to Raise Awareness of the knowledge-to-Action Gap in Information Ethics.**
  - Koji Tanaka, Honomi Miwa, Mitsuru Ikeda and Masahiro Hori
  - 955

- **An Empirical Study on the Influencing Factors of ICT Application: A Large Scale Survey for Middle and Primary Schools in China**
  - Wei Wang and Chun Lu
  - 965

- **PIVOTeeING: A Flipped Approach in a Postgraduate Solid State Devices Course**
  - Lakshmi T G, Soumya Narayana, Harshavardhan Penugonda, Dhirendra Vaidya, Vishvendra Poonia, Swaroop Ganguly and Sahana Murthy
  - 974

#### Short papers

- **Computational Thinking Development through Programmable Robotics Activities in STEM Education in Primary Schools**
  - Siu-Cheung Kong and Chan-Chio Lao
  - 984

- **Designing Mobile Applications for Improving Positive Behaviour for Learning (PB4L) Pedagogy**
  - Lou Reddy and Nilufar Baghaei
  - 990

- **Teachers’ Perception of IT in Science Education**
  - Der-Thang Chen and Wenjin Bo
  - 996
Surveying Indonesian Teachers’ Design Belief and TPACK for 21st Century Oriented Learning

Ching Sing Chai, Joyce Koh, Uma Natarajan, Pei-Shan Tsai, Murni Ramli and Ari Widodo 1001

Supporting Learning by Doing in a Work-process-oriented Curriculum

Yongwu Miao and Ulrich Hoppe 1007

Developing Pre-service Teachers’ 21st Century Teaching Competencies via Digital Storytelling

Wan-Lin Yang 1013

Current Situation of Chinese Primary and Secondary Innovative Teachers’ Evaluation in Maker Classes

Jun-Hao Shan and Shuai-Shuai Li 1018

**Posters**

Students’ In-Class Answering Activities on Facebook: Effects on Participation, Learning Satisfaction and Anxiety

Yu-Hsin Liu and Fu-Yun Yu 1024

Student-Generated Feedback for Online Student-Generated Multiple-Choice Questions: Effects on Question-Generation Performance and Perspective-Taking Development

Fu-Yun Yu and Wan-Shan Wu 1027

Training System for Puncture Operation Force Adjustment in Hemodialysis

Ren Kanehira, Kazunori Yamazaki and Hideo Fujimoto 1030

Mobile Learning in Higher Education in Sudan

Adam Tairab, Huang Ronghuai, Mohannad Taha and Kirk Perris 1033

Close Reading of Science Texts with Online Annotations

Guillaume Schiltz, Sarah Frederickx and Norman Sieroka 1036

A Secular Trend Analysis of the Effects of Using ICT in University Education

Yasuhiro Tsuji, Rieko Inaba, Mieko Takahira and Mana Taguchi 1039

**Early Career Research Paper**

The Role of a Technology and a Classroom Activity for Improving EFL Learners’ Oral Performance

Kae Nakaya 1042

**Author Index**

1044
Investigating the Effects of Cognitive and Metacognitive Scaffolding on Learners using a Learning by Teaching Environment

Cristina DUMDUMAYAab*, Michelle BANAWANC, Ma. Mercedes RODRIGOa, Amy OGANd, Evelyn YARZEBINSKi & Noburo MATSUDAe

aDepartment of Information Systems and Computer Science, Ateneo de Manila University, Philippines
bInstitute of Computing, University of Southeastern Philippines, Philippines
cComputer Studies Department, Ateneo de Davao University, Philippines
dHuman-Computer Interaction Institute, Carnegie Mellon University, USA
eCollege of Education and Human Development, Texas A&M University, USA
*cedumdumaya@usep.edu.ph

Abstract: We compared the effects of cognitive and metacognitive scaffolding on students’ performance within a learning-by-teaching intelligent tutor for algebra. Results revealed that metacognitive scaffolding facilitated learning for low prior ability learners while high prior ability students’ performance was obstructed by their refusal to follow the hints. Moreover, the study found that metacognitive actions related to self-regulatory learning’s reflection subfunction correlates negatively to the learning outcomes of low ability students.

Keywords: Intelligent tutoring systems, learning-by-teaching, scaffolding, self-regulated learning, SimStudent

1. Introduction

Scaffolding refers to the contingent and faded support aimed at improving students’ performance, which closely relates to the Zone of Proximal Development (ZPD) theory (Vygotsky, 1978). Scaffolding motivates students to learn complex ideas that are beyond their grasp (Jumaat and Tasir, 2015). Research has shown evidence that one’s learning can be improved through scaffolding in technology-enriched learning environments. Scaffolding triggers students to express knowledge and skills in various ways (e.g. feedback, transformation, self-explanation, and diagramming), that they will otherwise miss without prompts coming from a more enabled peer, which may result in missed opportunities to promote recall of prior knowledge, deeper understanding, and processing for knowledge integration (Demetriadis et al., 2008). Scaffolds can help learners improve their learning both at cognitive and metacognitive levels (Azevedo and Hadwin, 2005).

Cognitive scaffolding supports learning by directing students towards learning-appropriate goals such as focusing their attention on the task at hand (Demetriadis et al., 2008) and assisting them to overcome barriers in solving problems through modelling, providing hints, and coaching techniques (Ahern, 2009). Metacognitive scaffolding supports learning by focusing students’ awareness on their own cognition and on their understanding of the activities they are engaged in (Jumaat & Tasir, 2015). Through this approach, students could be motivated to plan strategies, evaluate and monitor learning process and performance (Efklides, 2008) based on their set goals, and adjust strategies to improve task effectiveness and their performance (Kinnebrew et al., 2014). Such executive functions are important components of self-regulated learning (SRL).

While older ITSs provide cognitive, domain-level support in forms of on-demand help and feedback (Roll, et al., 2006; Luckin and Du Boulay, 1999; Conati and VanLehn, 2000), the number of ITSs providing support for the activation of metacognitive skills have grown. Some examples of these ITSs are Geometry Cognitive Tutor (Schwonke et al., 2013), SimStudent (Matsuda et al., 2014), Betty’s Brain (Leelawong and Biswas, 2008) and SlideTutor (Feyzi-Behnagh et al., 2014). Despite extensive research in this field, the question on how these types of support affect students is still under investigation. We hope to contribute to the body of work on the relationship between
scaffolding and learner achievement. We investigate whether different types of support benefit different types of students in different ways and how learner usage patterns mediate these effects. In this paper, we investigate the effect of both cognitive and metacognitive supports on students using SimStudent, a learning-by-teaching ITS for algebra. Our research questions are:
1. What is the effect of cognitive and metacognitive scaffolding on learners' achievement?
2. How do cognitive and metacognitive scaffolding affect the achievement of learners with varying levels of prior knowledge?
3. Do hint and resource usage mediate learners' achievement?

2. Research Hypotheses

Scaffolding was shown to improve student performance in problem solving (Roll et al., 2006; Matsuda et al., 2014; Leelawong and Biswas, 2008). We hypothesize that students in both cognitive and metacognitive conditions will have increased achievement given these types of scaffolding.

Cuevas et al. (2002) found that cognitive and metacognitive scaffolding effects were strongest for low ability participants. Luckin and Hammerton (2002) and Weerasinghe and Mitrovic (2006) have shown that metacognitive scaffolding benefits low prior knowledge learners more than high prior knowledge learners. Thus, we hypothesize that lower prior ability students will benefit from scaffolding, with greater learning gains, among learner classes.

Previous work demonstrates that cognitive learning strategies (Berthold et al., 2007) and metacognitive prompts (Sonnenberg and Bannert, 2015; Jumaat and Tasir, 2015) mediated students’ learning outcomes. We hypothesize that students who make greater usage of cognitive and metacognitive scaffolds and resources benefit more than those who have lesser utilization of these affordances.

This paper contributes to the literature in that it provides us with evidence about how these different types of support might affect different students. Moreover, most research conducted in this area use data from western, English-speaking participants. Our study used a data set collected through ITS deployment in the Philippines, providing cross-cultural data that is not always available.

3. The Learning Environment

3.1 SimStudent

SimStudent is a virtual teachable agent that learns procedural skills inductively from the examples given by a human tutor. SimStudent attempts to solve the problem one step at a time, occasionally asking the human tutor about the correctness of each step or requesting for a demonstration. From the feedback and demonstrations, SimStudent generates production rules that represent the skills learned (Matsuda et al., 2014).

3.2 APLUS: Artificial Learning Environment using SimStudent

In order to use SimStudent as a teachable agent for peer tutoring, it is embedded into a game-like learning environment called APLUS (Artificial Peer Learning environment Using SimStudent). Figure 1 shows an annotated sample screenshot of APLUS. In this example, the student tutor teaches a customizable pedagogical agent, visualized with an avatar named Alex (“c” in Figure 1). In the tutoring interface (“b” in Figure 1), the tutor and Alex, the SimStudent, collaboratively solve problems. The tutor gives the problem 3x+6=15 for Alex to solve. Alex suggests a transformation and enters “add 3”. After performing a step, Alex verifies with the tutor if the step is correct. The tutor responds by clicking the feedback buttons (Yes/No) (“d” in Figure 1). If the tutor gets stuck or is preparing for tutoring, he/she can access the examples and other system resources by clicking the corresponding tab in the resources bar (“a” in Figure 1). The student tutor can measure how much Alex has learned by quizzing. Quizzes can be administered multiple times. Mr. Williams (“f” in Figure 1), the metatutor agent, presents the summary of the SimStudent’s performance on the quiz.
(“e” in Figure 1). SimStudent cannot proceed to the next level quiz unless all equations in the current level are correctly solved. The student tutor can also asks Mr. Williams for help when they encounter difficulties.

In this study, we used modified versions of this ITS to provide adaptive scaffolding in two ways: (1) Cognitive scaffolding provides adaptive assistance on how to solve equations. When the tutor asks Mr. Williams about the correctness tutee’s action (e.g. “Is this step correct?”), feedback is given (i.e. “No, subtract 10 would not be the right thing to do.”). When the tutor does not know the correct next step and asks for help (i.e. “What’s the next step?”), Mr. Williams answers by demonstrating the step (e.g. “What do you get when you apply the transformation 'subtract 5/9' to j+5/9? You need to enter j on the left side.”). (2) Metacognitive scaffolding provides adaptive assistance on how to teach or proceed with tutoring. When the tutor requests assistance (e.g. “What should I do now?”), Mr. Williams provides one of the four types of metacognitive help: (a) quiz assistance to suggest when students should take the quiz and why, (b) problem selection assistance to suggest what problem students should pose next and why, (c) resource assistance to suggest when students should review a particular resource and why, and (d) impasse recovery assistance to suggest a problem restart or give a new problem when students are stuck for a predetermined amount of time.

Aside from the adaptive scaffolding the system provides, APLUS is embedded with features supporting tutors in activating self-regulatory processes. These affordances aid tutors in setting their goals by learning what the task is all about, planning strategies by understanding the subject domain, teaching their tutee, and assessing their own understanding and performance. These features are described elsewhere (Matsuda et al., 2014, Matsuda et al., 2016).

4. Methodology

4.1 Structure of the Study

The study took place over five consecutive days and one day two weeks after the fifth day. On day 1, the students completed a 40-minute pre-test to assess their proficiency in solving equations and a pre-session survey to assess motivational attitudes towards learning algebra. On day 2, students were given instructions on how to use the software, after which they started tutoring their SimStudent. Tutoring sessions took place for three consecutive days. Each session lasted 40 minutes. Following the tutoring sessions, students were given a 40-minute post-test (Procedural Skill Test and Motivational Test) on day 5. A delayed post-test for Procedural Skill Test was administered two weeks after the fifth day.
4.2 Written Tests

The Procedural Skill Test contained 10 problems; 1 one-step equation, 3 two-step equations and 6 equations with variables on both sides. Three versions of isomorphic tests were randomly used for pre-, post-, and delayed-tests to counterbalance the test differences.

The Motivational test (MT) was used to assess students’ motivation in learning Algebra before and after the intervention. The pre-session MT consisted of 15 questions: 14 7-point Likert-style questions with response scale ranging from 1 (Not all true) to 7 (Very true), and 1 open-ended question to allow participants to explain what he/she thought was easy or hard about learning Algebra. The post-session MT consists of 17 questions, 16 7-point Likert-style questions and 1 open-ended question to explain what the participant thought was easy or hard about teaching SimStudent.

4.3 Interaction Logs

The system automatically recorded user-system interactions. The log file records specific parameters such as problems tutored, examples reviewed, hints requested, quiz attempts, and feedback provided (i.e., when the SimStudent asks the tutor for confirmation of an action (Yes/No) and the tutor’s explanation of a particular action). It also logs resource usage information, as well as requested (i.e. user-initiated) and proactive (i.e. metatutor-activated) scaffold information.

5. Results

The study was conducted in three high schools in the Philippines: two private schools (Davao City and Baguio City) and one public school (Quezon City). A total of 154 Grade 8 students voluntarily participated in the experiment. Each one was randomly assigned to one of the two versions of SimStudent: the experimental condition where metacognitive scaffolding was provided, and the control condition where cognitive scaffolding was provided. Out of those 154 students, 67 took the pre-, post-, and delayed post-tests and participated in all three days of tutoring sessions. Those 67 students were included in the following data analyses.

To investigate the impact of students’ prior knowledge on their learning, students were tertile-split (Sabourin et al., 2012) based on their pre-test scores into LPK=low prior knowledge-, APK=average prior knowledge-, and HPK=high prior knowledge- learners. Table 1 shows the number of students per condition, the gender and prior knowledge splits.

<table>
<thead>
<tr>
<th>Condition</th>
<th>HPK</th>
<th>APK</th>
<th>LPK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metacognitive(N=32, M=20, F=12)</td>
<td>8</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Cognitive(N=35, M=23, F=12)</td>
<td>8</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>

5.1 Test Scores

Table 2 shows the average (and standard deviation) of the students’ procedural skills test (PST) scores. We used repeated measures ANOVA with condition as between-subject variable and time (pre, post, and delayed) as within-subject variable to examine how the students’ test scores varied before and after the interventions and to verify if the effects of scaffolds differ by condition. The results revealed a simple effect of time ($F(2,130)=10.491$, $p<0.001$). The post-hoc analysis revealed that the students’ scores increased significantly from the pre-test to post-test ($t(66)=-4.223, p<0.001$) and pre-test to delayed post-test ($t(66)=-3.632, p = 0.001$). However, condition was not a main effect ($F(1,65)=0.046; p>0.05$) indicating no significant difference on the effects of the interventions on students’ test scores. Further, there was no statistically reliable interaction between time and condition ($F(2,130)=0.916, p>0.05$).
Table 2: Means (and standard deviations) for Procedural Skills Test (PST).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive (N=35)</td>
<td>4.29(3.77)</td>
<td>5.34(3.59)</td>
<td>5.09(3.98)</td>
</tr>
<tr>
<td>Metacognitive (N=32)</td>
<td>4.63(4.16)</td>
<td>5.22(3.96)</td>
<td>5.47(4.17)</td>
</tr>
<tr>
<td>Aggregate (N=67)</td>
<td>4.45(3.94)</td>
<td>5.28(3.74)</td>
<td>5.27(4.05)</td>
</tr>
</tbody>
</table>

5.2 Effects of Scaffolding on Learning Gains

A between-subject comparison of pre-test scores showed no significant difference in prior knowledge among students in the cognitive and metacognitive groups \((t(65)= 0.350, p>0.05)\), tested using a two-tailed t-test.

We examined whether metacognitive and cognitive scaffolding helped certain students differently by running repeated measures ANOVA for each condition, with time (pre-, post-, delayed post) as within-subject variable and prior knowledge grouping (HPK, APK, LPK) as between-subject variable. Table 3 shows the PST scores and normalized learning gains for each intervention type, grouped by prior knowledge.

Table 3: Procedural Skills Test (PST) scores and normalized gains, by prior knowledge group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed Posttest</th>
<th>Normalized Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metacognitive</td>
<td>HPK 9.63(0.52)</td>
<td>9.5(1.07)</td>
<td>9.5(0.76)</td>
<td>-0.0014(0.009)</td>
</tr>
<tr>
<td></td>
<td>APK 5.46(2.73)</td>
<td>6.15(2.97)</td>
<td>6.69(3.40)</td>
<td>0.0073(0.017)</td>
</tr>
<tr>
<td></td>
<td>LPK 0(0)</td>
<td>1.0(1.34)</td>
<td>1.09(1.92)</td>
<td>0.01(0.0134)</td>
</tr>
<tr>
<td>Cognitive</td>
<td>HPK 9.5(0.76)</td>
<td>9.5(0.76)</td>
<td>9.63(0.52)</td>
<td>0.1875(0.372)</td>
</tr>
<tr>
<td></td>
<td>APK 5.31(1.60)</td>
<td>6.23(2.95)</td>
<td>6.54(2.79)</td>
<td>0.2833(0.492)</td>
</tr>
<tr>
<td></td>
<td>LPK 0.36(0.50)</td>
<td>2.14(1.70)</td>
<td>1.14(1.61)</td>
<td>0.1825(0.180)</td>
</tr>
</tbody>
</table>

The test revealed a main effect of prior knowledge; \(F(2,29)=51.178, p<0.001\) for the metacognitive group. Time was also a main effect: \(F(2,58)=3.337, p=0.042\); however, no statistically reliable interaction between time and prior knowledge was observed; \(F(4,58)=1.149, p>0.05\). Comparably, learners in the cognitive group also revealed a main effect of time; \(F(2,64)=5.111, p=0.009\) and prior knowledge; \(F(4,32)=82.944, p<0.001\), but revealed no statistically significant interaction between time and prior knowledge; \(F(4,64)=2.037, p>0.05\). These results suggest that the effects of both cognitive and metacognitive scaffolding do not vary with learners’ prior knowledge.

Table 4: Comparison of normalized gains.

<table>
<thead>
<tr>
<th>Group</th>
<th>HPK vs. LPK</th>
<th>HPK vs. APK</th>
<th>LPK vs. APK</th>
<th>LPK vs. APK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metacognitive</td>
<td>t(17) = -2.194</td>
<td>p=0.042</td>
<td>t(19) = -1.321</td>
<td>p&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>t(22) = 0.432</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>t(19) = 0.042</td>
<td>p&gt;0.05</td>
<td>t(19) = -0.472</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t(25) = -0.716</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However, when the learners’ normalized learning gains were analyzed to account for the variance in learner’s prior knowledge, the results revealed that LPK learners given metacognitive scaffolding posted a higher pre- to post-test learning gains than HPK learners; \(t(17)= -2.194, p=0.042, d=0.99\). A similar analysis revealed no significant difference for the learning gains of HPK and LPK students of the cognitive group; \(t(19) = 0.042, p>0.05\). Normalized learning gain was computed using the standard formula: \(\text{normalized gain} = (\text{posttest-pretest})/(\text{maximum possible score-pretest})\). Comparisons of normalized gains between HPK and APK, and LPK and APK.
learners, for both groups, revealed no significant differences (see Table 4). Moreover, the analysis revealed that metacognitive group’s HPK learners did not gain learning from the intervention.

5.3 Hint Utilization

To explore on what might cause LPK learners to gain more from metacognitive scaffolding than HPK learners, we compared the learners’ hint utilization. Overall, learners given metacognitive scaffolding exhibited a low hint utilization score, only 14% of the given hints were utilized by the students. Comparison of HPK and LPK learners’ metacognitive hint utilization $(\text{FollowedHints})$ revealed no reliable difference; $\text{FollowedHints}_{MC_H, \text{AVE(SD)}} = 0.8(0.7) \text{ vs } \text{FollowedHints}_{MC_L, \text{AVE(SD)}} =1.0(0.6), t(17)= -0.81, p>0.05$. Moreover, no significant difference was observed in their non-utilization of hints $(\text{IgnoredHints})$; $\text{IgnoredHints}_{C_H, \text{AVE(SD)}} = 3.13(2.23) \text{ vs } \text{IgnoredHints}_{C_L, \text{AVE(SD)}} =4.18(4.02), t(17)=-0.669, p>0.05$.

Further, we investigated whether utilization and non-utilization of metacognitive scaffolds predict the post-test scores of learners in each ability grouping. We ran centered regression analysis, independently, for HPK and LPK groups. The results revealed that following hints, $\text{FollowedHints}(p>0.05)$, had no correlation to HPK learners’ posttest (PT) scores while ignoring hints, $\text{IgnoredHints}(p=0.004)$, negatively correlates to their posttest scores. The model equation is $\text{PT} = 8.063 -0.521\text{FollowedHints} -0.429\text{IgnoredHints} +0.668C_{\text{mean}} ; F(3,4)=14.335, p=0.013, r^2=0.915$. The derived model for the group’s LPK learners was not significant: $F(2,8)=3.314, p>0.05$.

5.4 Resource Utilization

Did the high prior and low prior learners use the resources differently? Does usage mediate learning outcomes? To answer these questions, we performed frequency and regression analyses on both user-initiated and metatutor-prompted resources’ usage. Table 5 shows the average (and standard deviations) of HPK and LPK learners’ resources’ access frequency (RA) and usage duration (DUR).

<table>
<thead>
<tr>
<th>Prior Knowledge Group</th>
<th>Resource Access, RA</th>
<th>Duration, DUR (in sec)</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPK</td>
<td>6.25(7.086)</td>
<td>427.96(313.98)</td>
<td>$F(3,4)=4.485, p&gt;0.05$</td>
</tr>
<tr>
<td>LPK</td>
<td>13.27(11.27)</td>
<td>588.80(523.43)</td>
<td>$F(2,8)=0.278, p&gt;0.05$</td>
</tr>
</tbody>
</table>

The comparison between HPK and LPK learners’ RA and DUR revealed no significant difference, $t(17)=-1.547, p>0.05$, and $t(17)=-0.771, p>0.05$, respectively. A regression analysis revealed that both RA and DUR were not predictors of the posttest scores of either HPK or LPK learners, when pretest is controlled.

Table 6. Frequency, means and standard deviations of Metacognitive-coded events.

<table>
<thead>
<tr>
<th>Metacognitive Events</th>
<th>HPK (N=8)</th>
<th>LPK (N=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Mean(SD)</td>
</tr>
<tr>
<td>SetGoal</td>
<td>2</td>
<td>0.25(0.71)</td>
</tr>
<tr>
<td>Plan</td>
<td>11</td>
<td>1.38(2.77)</td>
</tr>
<tr>
<td>SeekInfo</td>
<td>4</td>
<td>0.50(0.76)</td>
</tr>
<tr>
<td>ReadInfo</td>
<td>69</td>
<td>8.63(4.21)</td>
</tr>
<tr>
<td>Monitor</td>
<td>18</td>
<td>2.25(1.58)</td>
</tr>
<tr>
<td>Evaluate</td>
<td>24</td>
<td>3.00(2.00)</td>
</tr>
<tr>
<td>Total</td>
<td>128</td>
<td>16.50(7.56)</td>
</tr>
</tbody>
</table>

We further examined whether activation of certain metacognitive events can be linked to students’ achievement. We mapped and coded students’ behaviors based on the integrated cognitive and metacognitive model for learning (Segedy et al., 2011) and analyzed the learning activities of
the students. We took the frequencies of both user-initiated and metatutor-prompted metacognitive activities. Table 6 shows the frequencies and means (and standard deviations) of the metacognitive events.

A comparison of means showed no significant difference in the number of metacognitive events used by HPK and LPK learners, \((t(17)=-0.61, p>0.05)\). However, HPK learners showed a high count on ReadInfo activities while LPK learners demonstrated high frequencies on both ReadInfo and Evaluate events.

Regression analysis was used to investigate whether the observed effect of the intervention on the learners’ post-test scores was influenced by these metacognitive actions. The test revealed that ReadInfo\((p=0.012)\), Monitor\((p=0.043)\) and Evaluate\((p=0.003)\) significantly correlate to the learning gains of LPK learners. These variables were regarded as possible mediators for pre-posttest learning gains of low ability students. The model equation is \(PT = 1.83 + 0.55\text{ReadInfo} - 0.48\text{SeekInfo} - 0.75\text{Monitor} - 0.57\text{Evaluate}\), \(F(5,5)=8.579, r^2=.896, p=0.017\). We failed to find correlates to the learning gains of high ability students; \(F(6,1)=1.190, p>0.05\).

6. Discussion and Conclusion

In this study, we extend existing scaffolding research by examining how both cognitive and metacognitive supports, in a learning by teaching ITS for algebra, benefit different types of students, and how learner usage patterns mediate these effects. The data show that both cognitive and metacognitive scaffolding increase learners’ achievement. This finding supports the effectiveness of the learning by teaching paradigm in improving students’ proficiency in problem solving (Matsuda et al., 2010, Matsuda et al., 2013).

The data also show that metacognitive scaffolding was more beneficial for low prior ability than for high prior ability learners as shown by higher learning gains, which corroborates previous results (Roll et al., 2014). Prior literature has established the importance of prior knowledge for a meaningful and effective learning by teaching exercise. The metacognitive scaffolds compensate for a learner’s lack of relevant domain knowledge (Chen & Chou, 2016). Further, metacognitive scaffolding enables students to reflect on the processes involved in problem solving, making students’ learning experience more memorable and coherent.

Scaffolding assists students in their learning process. Those who exploit the system’s suggestions and assertions approach their tasks systematically and have learned the importance of information seeking activities for a successful learning engagement (Segedy et al., 2013). Ironically, the learners in this study ignored the metatutor’s hints, which may have caused the negative learning gains of high ability learners. We hypothesize that the hint avoidance behavior of students could be because of low goal or task commitment, hence they exerted lesser effort in the task by avoiding hints (Linderman et al., 2003). Commitment to set goals is an essential attribute of self-regulated learning (Boekaerts, 1996). Goal commitment directly influenced both students’ intention to persist in a task and their actual persistence behavior (Allen and Nora, 1995). The challenge, therefore, is to help students develop not only their cognitive abilities but also their behavioral attributes (e.g. persistence) by helping them become self-regulated learners.

In this study, we found that metacognitive activities ReadInfo, Evaluate and Monitor relate to low prior ability students’ learning performance. ReadInfo, Evaluate and Monitor are metacognitive actions that support SRL’s process of reflection (Siadaty et al, 2016). ReadInfo, which pertains to viewing quiz problems’ solution, is a metacognitive activity to evaluate one’s learning process and compare one’s work with others. Besides enacting ReadInfo, low prior knowledge learners were also performing actions which allowed them to Monitor and Evaluate their understanding of the task and the processes they execute to complete the task (Siadaty et al, 2016). ReadInfo seemed helpful to students in achieving better performance; however, there were metacognitive events that failed to help low ability students. Actions related to monitoring and evaluating results and processes appeared to be more harmful than helpful in facilitating low ability learners. Monitor (e.g. self-explanation), which probes a tutor’s reasoning (Kinnebrew and Biswas, 2014), appeared to hurt student learning. In comparison with ReadInfo, which is instructive in nature and shows the student what to do, Monitor is reflective. Monitoring requires students to think about
their own knowledge and use it to help themselves out of their intellectual rut. If that knowledge is weak in the first place, they may not be able to help themselves. Rather, what they might need is an external intervention to help them move forward. The current software version does not engage students in a dialog to help correct misconceptions; hence, students have no opportunity to learn that their explanation is wrong or to construct deeper knowledge on the subject matter. Thus, there is a need for a constructive feedback mechanism to make self-explanation effective. Moreover, Evaluate, which is mapped to the system’s Quiz feature, obstructed student learning. We suspect that frequent quizzing disengaged students from the tasks by dissuading students from working on the problem and utilizing the features that could prepare them for the task, a behavior which mirrors “gaming the system” (Baker et al., 2008).

The contingent relationship of metacognition and self-regulation and their impact on student learning (Wagster et al., 2007; Kinnebrew et al., 2014; Duffy and Azevedo, 2015) and task persistence (Pintrich et al., 2000) have been demonstrated in previous studies. Self-regulated learners have the capability to evaluate their tasks, reflect on their performance, and manage their own learning and behavior. They feel empowered to succeed, thus they utilize strategic approaches to complete the tasks successfully and to sustain their efforts when difficulties are encountered. Metacognitive skills activate students’ self-regulation, which in turn helps in developing their intellectual curiosity and persistence (Paris and Winograd, 1990). In contrast, the lack of metacognitive abilities may result in task disengagement and a lack of sustained effort to complete the task.

In summary, the findings of this study corroborate previous findings on the effect of tutor learning in the context of learning by teaching as indicated by the learners’ improved post test scores. We also found that students’ hint avoidance behavior impedes their performance. Furthermore, we found that metacognitive scaffolding led to better learning gains for low prior ability learners. Their performance is mediated by metacognitive activities (i.e. ReadInfo, Evaluate and Monitor), however, metacognitive activities which are related to the reflective process or to the self-monitoring sub-function of self-regulation negatively affect low ability learners’ performance.

7. Limitations and Future Work

One of the limitations of this study was its small sample size (Metacognitive \(N = 32\) and Cognitive \(N=35\)). When grouped according to prior domain abilities, by condition, sample size was reduced further, which limits the generalizability of the findings of this study. The small sample size was due to the high attrition of participants, and the long duration of the study (spanning almost 3 weeks). Also, our data collection coincided with the performance evaluation week of some deployment sites. Given that participation was voluntary, participants gave priority to classroom performance tasks over the experiment, resulting in high attrition. Additionally, we did not collect affect or eye-tracking data in this study. This would have provided evidence for participants’ engagement during the ITS use, which would help in making more accurate inferences about the learners’ cognitive and metacognitive processes. With no such evidence, we could not confirm whether the participants really paid attention to the resources provided to prepare them for their task as they were utilizing it.

Recent work has emphasized the interdependence of non-cognitive (e.g. critical thinking, self-regulation, persistence) and cognitive skills to prepare students for the increasingly complex nature of 21st century jobs. Non-cognitive attributes support cognitive development. Unless closer attention is given to non-cognitive skills, cognitive skills’ development may fail. As part of our future work, the authors will conduct further studies to gain a better understanding of how metacognitive learning events affect students’ patterns of thoughts and behavior. Action modeling is one potential technique to quantify and detect incidence of non-cognitive factors such as persistence from the interaction logs.

Acknowledgements

The research reported here was supported by National Science Foundation Awards No. 1643185/1252440. The authors would like to thank Ateneo Laboratory for the Learning Sciences,
Ateneo de Manila University, the Ateneo de Davao University, Krus Na Ligas High School, and the University of the Cordilleras for the support they extended to this study.

References


Learning Arithmetic Word Problem Structure with a Picture Combination Application in Kindergarten

Pedro Gabriel Fonteles FURTADO*, Tsukasa HIRASHIMA, Yusuke HAYASHI & Kazumasa Maeda

*Hiroshima University, Japan
bInternational Pacific University, Japan
*gabrielpff2@gmail.com

Abstract: Teaching number sense to kids before elementary school has been researched thoroughly, but there has been a lack of research into teaching the structure of word problems to kids of the same age, despite of their classroom popularity and the trouble students have with them. Our research focuses on creating a textless, image rich application empowered by the Triplet Structure Model. This model has been used with success to teach the conceptual structures of word problems, by allowing young kids to interact with the structure of arithmetic word problems. An experiment has also been done in a local kindergarten. Experiment results show that young kids can also interact successfully with the model.

Keywords: Pre-math, preschool, word problems, arithmetic

1. Introduction

Past research has shown that math proficiency in the early years, such as number sense, can predict math performance in school life (Jordan, 2010). This has been verified all the way to the third year of elementary school. This suggests that building math skills before elementary school can greatly impact how the student deals with math in the coming years. Interventions to teach number sense for both kindergarten and pre-kindergarten students have been done before (Griffin, 2004; Dyson, Jordan, & Glutting, 2013; Starkey, Klein, & Wakeley, 2004; Wilson, Dehaene, Dubois, & Fayol, 2009). These studies focus on whole number understanding. However, these interventions do not focus on word problems which is a big part of how students interact with math after kindergarten. Students will usually encounter word problems in early arithmetic classes in their first year of elementary school. Those word problems describe an addition or a subtraction. Students are often required to:

1. Write the arithmetic expression corresponding to a certain word problem;
2. Calculate the answer by solving the arithmetic expression.

Focusing on the first step, (Rivera, 2014) states that elementary word problems need to be conceptually understood. It states that teachers often teach children to rely on keywords in order to transition to the arithmetic expression. It also states that this keyword-based method shows a lack of conceptual understanding that is necessary for solving harder problems. An analysis by (Hegarty, Mayer, & Monk, 1995) also points out this issue, stating that students that rely on keywords instead of creating a mental model of the problem are usually unsuccessful. Students need to be able to understand the roles of the 3 quantities inside the story. But how to help students develop this conceptual understanding and build the necessary mental model of the problem?

The Triplet Structure Model that will be introduced below attempts to offer a solution by creating a 3 sentence representation bridging the word problem and its arithmetic expression (Hirashima, Yamamoto, & Hayashi, 2014). Monsakun is an interactive learning environment for learning by problem-posing of arithmetic word problems designed based on the triplet structure model. In Monsakun, a learner poses an arithmetic word problem by arranging three sentences cards by selecting three cards from multiple provided sentence cards. Monsakun diagnoses the posed problem based on the triplet structure model and gives feedback. It has been used in elementary schools and it has shown promising results so far, with some experiments suggesting that Monsakun use results in better problem-
solving skills, for examples, in forward-thinking addition and subtraction problems (Hirashima, Yokoyama, Okamoto, & Takeuchi, 2007), in reverse thinking addition and subtraction problems (Hirashima & Kurayama, 2011; Yamamoto, Kanbe, Yoshida, Maeda, & Hirashima, 2012) or in multiplication and division problems (Yamamoto, et al., 2014). The problem-posing process has also been analyzed for addition and subtraction problems (Supianto, Hayashi, & Hirashima, 2016; Hasanah, Hayashi, & Hirashima, 2017). These researches reported that affective responses were also positive. Those results suggest that the Triplet Structure Model can be used in the classroom as a way to deepen this conceptual understanding.

Kindergarten students may also be able to learn from the Triplet Structure Model. There is a problem though. Monsakun is based on written sentences. Thus, younger students that aren't used to reading are not in a position to take advantage of Monsakun. So in order to verify if these young students can successfully interact with the model or not, a new application must be developed. This new application must not rely on text in order to provide this interaction.

As for the paper structure, in section 2 we’ll clearly state our research questions. In section 3 we’ll introduce the Triplet Structure Model. On section 4 we’ll introduce our application design and show how it relates to our research questions and to the Triplet Structure Model. In section 5 we show our experiment design and its results. Lastly, on section 6, we conclude our paper by relating our results to our research questions and expanding on further impacts of the results.

2. Research Questions

Our main research question is "Can kindergarten students meaningfully interact with the Triplet Structure model?" In order to answer this, the following questions must be answered:

1. How to design an application for interaction between the kindergarten students and the Triplet Structure Model?
2. Can they use this application? What is their response to the application?
3. Is their understanding and use of the application meaningful and useful? Or are they aimlessly interacting?

To answer the first question, we will show the design of our application and how it relates to the model. To answer the second question, we will run experiments with kindergarten classes and examine their initial response. Afterwards, examining their performance through the application log data will help us answer the third question.

3. The Triplet Structure Model

The triplet structure model is a model that binds the 3 quantities of one arithmetical operation (operand, operant and result quantities) to contextual story roles by using 3 quantity sentences. This is illustrated on Figure 1. It is usually used to describe arithmetic word problems where one of the quantities is unknown.

![Figure 1 Triplet Structure Model](image)
These quantity sentences can be classified into two types: independent quantity and relative quantity sentences. Independent quantity sentences state the existence of a certain number of objects, such as in the case of "there are two apples". Relative quantity sentences, such as "two apples are eaten" depend on the previous existence of a certain number of objects (two apples cannot have been eaten if apples aren't there), which means that they depend on independent quantity sentences to state that existence.

![Increase and composition addition stories](image)

The example on Figure 1 contains a case of subtraction, more specifically what the Triplet Structure Model calls a "decrease story". Figure 2 has examples of composition and increase stories, both of those classified as addition stories. Both increase and decrease stories are composed of two independent quantities intercalated by a relative quantity. The relative quantity serves to show how much the amount of the object changed (increased or decreased, depending on the story), while the independent quantities show the amount of objects before and after the change. All 3 quantities in increase and decrease stories must refer to the same object, or else the story is invalid ("there are two apples, one apple is eaten, there is one banana" is not valid, for example).

Composition stories have a slightly different structure. On composition stories the relative quantity comes in the end, with two independent quantities coming before it. The two independent quantities must then refer to different types of objects, while the relative quantity describes the total number of the two objects together. On decrease and increase stories, the relative quantity created a change in the amount of a certain object. In the composition case, the relative quantity it is making a numeric observation about the previous defined quantities without changing their value. Different objects here could be "John's apples" and "Mary's apples", while the relative quantity would be ("John and Mary's apples put together").

There is one more story type that is out of the scope of this paper. It’s the comparison story. Due to the nature of the story, we found it difficult to show it using pictures and decided to not include it in our design. We might include it in future research once we find a satisfactory way of showing it. More information on it and on the Triplet Structure Model can be found in (Hirashima, Yamamoto, & Hayashi, 2014).

4. Application Design

This application must allow for interactivity with the Triplet Structure Model without relying on text. Our solution has been to use pictures and spoken sound. There are two types of pictures. The first one is horizontally large and called overall story pictures. They show the entire problem at once. The second type is the story piece picture. They are small, squared and represent each sentence in the Triplet Structure Model sentence. Understanding the relationship between the overall story pictures and the story piece pictures is similar to understanding how a problem is made up in the Triplet Structure Model. The design of the pictures will be further introduced below. In the application we also ask for users to connect the pictures to numbers. This connection brings them a step further to connect the in-context parts of a problem to the out-of-context parts of an arithmetic expression. This type of connection is in-line with the conceptual understanding of word problems described in previous sections.

We have divided the application activities into levels. Each level has been described below. Level 3, in particular, is critical to the application. It focuses on connecting story piece pictures to big pictures. Performing this requires students to be able to divide the big picture into 3 small parts, each related to a meaningful quantity.
Also worth noting is Level 6 connects the pictures to numbers. At first we ask students to connect story piece pictures to numbers. Since each small picture refers to 1 quantity, connecting 1 meaningful quantity to 1 number is not a hard task. But later the application requests users to connect 1 big picture to 3 numbers. Since no other help is given to the user, he has no choice but to visualize the 3 numbers related to the big picture. This is similar to writing the arithmetic expression of a given problem, but given only the picture of the problem. Students that are able to perform this task well should be apt to meaningfully connecting word problems to their arithmetic expressions.

All of the levels will be explained further below.

4.1 Image and Sound Design in Connection to the Triplet Structure Model

The two picture representations of a Triplet Structure Model story can be seen in Figure 3. Each picture also has an accompanying sound. On the left we have the overall story, where one picture contains all information to describe a story. On the right, we have the story pieces, in which 3 pictures are put together to describe a story.

While on the model we would have "There are 3 apples", in the picture sound we would have "there are 3 apples in the shelf". We describe the place the objects are in to make the connection between what the sound is saying and what the picture is showing stronger.

Story piece pictures come in two types, independent and relative pictures. They are correspondent to independent and relative quantities in the Triplet Structure Model. Independent pictures are usually composed of stationary objects. Relative pictures will usually describe some sort of action, like a boy inserting objects somewhere or an animal entering a place. By describing an action or movement, we can convey the same idea as the relative quantities of the Triplet Structure Model. We can be confident in the children’s interpretations of the pictures because very similar designs are used in Japanese textbooks, which are used in Japanese schools.

Overall story pictures represent the entire problem in a single picture. In the overall story picture in Figure 3, the three numbers of the problem can be seen, as
1. the number of watering cans in the shelf;
2. the number of cans the boy puts in the shelf;
3. The total number of cans after the boy inserts them.

So these 3 numbers are mapped to the 3 story piece pictures, creating the relationship between the overall picture and the story pictures. The sound related to the overall picture is a simple combination of the sounds of the 3 story pieces put together. While connecting the overall story to the 3 pictures may seem like a trivial task, it is not so simple. While the first two numbers are quite clear in the overall story picture, the third number, which represents how many cans there will be in the shelf, requires the student to understand the described story and then recognize that there will be two watering cans in the shelf after the boy is done. This number could be calculated by counting or by mental addition, it doesn’t matter. What matters is whether or not the student can interpret the story. Since not all quantities are explicitly shown, connecting the overall story picture to the 3 story piece pictures requires more than simply looking at the photo and trying to match the objects or scenery.

4.2 Level 1, Level 2-1 and Level 2-2

On Level 1, participants listen to audio describing a picture and then have to choose, from 3 pictures, which picture corresponds to the audio. In this level, the pictures are story piece pictures and not overall
story pictures. This is an introductory stage to introduce the pictures that can make up a story and their corresponding description. Level 2 focuses on connecting overall story pictures to their spoken narration, to ease students into understanding the content of the pictures. It is made up of two parts. Part 2-1 is based on true or false problems, participants hear the spoken narration and are shown one picture. They have to decide if the picture described in the narration is the picture being shown or not. Part two is similar to Level 1, where participants are shown 3 pictures and audio, having to decide which picture corresponds to the audio. Level 1 and 2 work together to introduce the pictures to the user.

4.3 Level 3-1

Level 3 focuses on connecting overall stories pictures to their story piece pictures and it’s made up of two parts. On part 3-1, we have true or false problems, with participants being shown one overall story big picture and 3 story piece pictures and then have to decide if the 3 story piece pictures correspond to the same story being shown in the big picture or not, by choosing from true or false buttons.

4.4 Level 3-2

On part 3-2, participants are shown an overall story picture and given multiple story piece pictures. They are asked to use 3 of the story piece pictures to make up a single story. The made up story must correspond to the same story being shown in the overall story picture. Figure 4 shows a screenshot of this setup. Participants are given 5 pictures. Since only 3 pictures make up a story, the remaining two pictures are dummy pictures. Dummy pictures are added to give students more to think about. Problems in Level 3-2 don’t all have the same difficulty, the number of blank spaces vary like this:

1. 1 blank: Problem 1 to 5. There is only one blank space. The other two spaces are automatically filled for the player. The player can move 3 of 5 cards.
2. 2 blanks: Problem 6 and 7. Two blank spaces. One space is automatically filled for the player. The player can move 4 of 5 cards.
3. 3 blanks: Problem 8 to 11. All blank spaces are fully opened. The player can move all 5 cards.

This difficulty progression is made up to more easily ease the participants in the workings of how Level 3 and constructing a story from its 3 pieces work. We stress once again that this is a key skill in the context of the Triplet Structure Model.

4.5 Level 4

Level 4 has participants listen to a narration of a story and then they are tasked with forming the story by using story pieces, similar to the second part of Level 3, with the difference being that in Level 3 it was a correspondence with the overall story picture, while on Level 4 it's with spoken narration. It is composed of 8 problems, with the first 3 being easier, only allowing users to move 3 pictures of 5, while the rest of the problems allow the user to move all 5 pictures.

4.6 Level 5

Level 5 shows participants an overall story picture and asks if that story belongs to a certain story type (the types being "increase", "decrease" and "combine"), with the participant having to choose "true" or
"false". This Level relates to how well the participant understands the concepts of increasing, decreasing and combining. It also relates to how the participants comprehend the story being show in each picture.

4.7 Level 6

In this Level we have participants connecting numbers to pictures. It is divided into two parts. In part one, students are tasked with connecting numbers to story piece pictures. In part two, students are asked to connect 3 numbers to a single overall story picture, a setup that can be seen in Figure 5. Like stated before, this is a problem that requires deep understanding of the three quantities that can be interpreted from a single picture. Users that are able to do this should be able to construct mental models that allow them to be successful problem solvers.

![Figure 5 Level 6 problem](image)

5. Experiments

5.1 Experiment Description

The participants of the experiment are 90 Japanese Kindergarten students (around 5 years old), divided into 3 classes averaging 30 students each. A math teacher briefly introduced the application for around 10 minutes, showing the first few problems to the participants by using a projector connected to a tablet. The participants were encouraged to give their opinion on the answer while the teacher advanced through the problems. Afterwards, students had around 20 minutes to interact with the application by themselves, by using an android tablet that contained the application and a headphone. A picture of a kindergarten student interacting with the application can be seen on Figure 6.

![Figure 6 A participant interacting with the application](image)

Afterwards, user log data was extracted from the tablets for analysis. Participants measured performance will be compared to their performance if they were “gaming the system”. Gaming the system (GTS), in this case, refers to the behavior displayed when students attempt to systematically take advantage of the way the system is made by randomly inputting answers (Baker, et al., 2008). To calculate this performance, we use the probability that students will solve a level in their first try. The probability that students will solve a level on their first try depends on both the level and the difficulty of the problem.
For example, on Level 4, the probability of solving the first three easy problems in one try is 1/3 (only one option with 3 choices). The probability of solving the last 5 problems is 1/60 (5 choices on the first card, 4 choices on the second card, and 3 choices on the third card). “(3 * 1/3 + 5 * 1/60) / 8” gives us the probability of solving every problem in one try on average. We can use a similar logic to calculate the average number of attempts. However, the calculated values are based on which problems the students have done on each level. While uncommon, students could stop midway through a level, restart and complete the level. This means that the calculated values are based on which problems the participants have done for this particular experiment.

A pre-test and post-test was not included in our experiment due to constraints of time in the school’s schedule.

Due to a poor choice in the application’s implementation, log data was only collected when the user completed a level. The analysis will still be done with the remaining data of the students who completed the levels. This means that our analysis will not include some data from levels that were stopped midway through the level due to time constraints.

5.2 Results and Discussion

Students from all classes were excited about using the application, describing it as fun and stating they wanted to use it more. Students were quick to progress through level 1 and 2 and found difficulty with level 3. We estimate that around 80 students started level 3 and only 13 students have completed it, due to time constraints. This restricts our analysis of later levels to the students who progressed quickly through the system.

Table 1. One sample t-test results comparing measured number of attempts per problem to the equivalent “gaming the system” calculated value. sd stands for standard deviation.

<table>
<thead>
<tr>
<th>Level</th>
<th>Average N. Attempts</th>
<th>Average N. Attempts (GTS)</th>
<th>T</th>
<th>df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.16 (sd=0.45)</td>
<td>2.00</td>
<td>-61.17</td>
<td>1081.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2-1</td>
<td>1.17 (sd=0.45)</td>
<td>1.50</td>
<td>-17.11</td>
<td>549.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2-2</td>
<td>1.24 (sd=0.54)</td>
<td>2.00</td>
<td>-32.82</td>
<td>533.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3-1</td>
<td>1.35 (sd=0.50)</td>
<td>1.50</td>
<td>-2.78</td>
<td>92.00</td>
<td>0.007</td>
</tr>
<tr>
<td>3-2</td>
<td>4.77 (sd=9.44)</td>
<td>13.18</td>
<td>-11.05</td>
<td>153.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>1.54 (sd=1.20)</td>
<td>19.64</td>
<td>-154.41</td>
<td>104.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>1.23 (sd=0.42)</td>
<td>1.50</td>
<td>-4.27</td>
<td>43.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6</td>
<td>1.31 (sd=1.01)</td>
<td>8.25</td>
<td>-27.35</td>
<td>15.00</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 2. One sample t-test results comparing measured ratio of problems solved in 1 try to the equivalent “gaming the system” calculated value. Data was constant for level 6 so the test was not performed.

<table>
<thead>
<tr>
<th>Level</th>
<th>Ratio of Solved in 1 try</th>
<th>Ratio of solved in 1 try (GTS)</th>
<th>T</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.87 (sd=0.16)</td>
<td>0.33</td>
<td>33.13</td>
<td>99.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2-1</td>
<td>0.84 (sd=0.18)</td>
<td>0.50</td>
<td>18.55</td>
<td>89.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2-2</td>
<td>0.82 (sd=0.18)</td>
<td>0.33</td>
<td>26.11</td>
<td>88.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3-1</td>
<td>0.75 (sd=0.27)</td>
<td>0.50</td>
<td>3.99</td>
<td>18.00</td>
<td>0.001</td>
</tr>
<tr>
<td>3-2</td>
<td>0.56 (sd=0.27)</td>
<td>0.21</td>
<td>4.82</td>
<td>13.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.81 (sd=0.24)</td>
<td>0.17</td>
<td>9.43</td>
<td>12.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>0.81 (sd=0.18)</td>
<td>0.50</td>
<td>5.38</td>
<td>8.00</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.88 (sd=0.00)</td>
<td>0.10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1 and Table 2 compares participants’ performance to their calculated GTS counterparts. Measured number of attempts is lower than the number of attempts by gaming the system (GTS). The ratio of problems solved in 1 try is much higher than the equivalent GTS value too. This shows that the strategies students employed to solve each level were more effective than GTS. This suggests that students were not blindly progressing through the application without thinking. Furthermore, this could be indicative that students are displaying the necessary skills to solve each level. It also shows signs that our pictures and their audio description fit with the participants’ interpretations of the pictures or else they would not be able to get high scores on the first two levels.
Figure 7. Performance on Level 3-2 problems. The bigger the symbol, the harder the problem. Problem 1 to 5 are the easiest. Problem 6 and 7 are harder. Problem 8 to 11 are the hardest.

Participants’ performance on Level 3-2 can be seen in Figure 7. The image shows the number of attempts divided by the calculated number of attempts of each problem during GTS. We divided the value to account for the difference in difficulty between the problems. In the graph, the bigger the symbol, the harder the problem. We can see that participants are clearly becoming more proficient in solving the problems as they advance through the level given the descending tendency in the graph. The spike on problem 6 is when the difficulty first raises. This spike can be attributed to the participants trying to adapt and understand what the task is asking them to do. After problem 6, their overall performance continues to increase, even though difficulty does not go down. The spike on problem 8, when the difficulty goes up again, isn’t nearly as sharp as the spike on problem 6, suggesting that after adapting to problem 6 students get a good grasp of what the activity is about.

Let’s now look at the problem types separately. Problems 8 and 11 are both “decrease type” problems with the same difficulty, which is the hardest. But the average went from above 15 mistakes to around 5 mistakes. For problems 6 and 9, which are “composition type” problems, the effect is easier to see. Problem 6 (difficulty 2) is easier than 9 (difficulty 3). Despite this difference, students made less mistakes on problem 9. This suggests that past experiences helped the students become better solvers.

Table 3. Comparison between participants who completed Level 6 and the entire group.

<table>
<thead>
<tr>
<th>Level</th>
<th>L6 kids average N. Attempts</th>
<th>Group average N. Attempts</th>
<th>L6 kids ratio of problems solved in 1 try</th>
<th>Group ratio of problems solved in 1 try</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00 (sd=0.00)</td>
<td>1.16 (sd=0.45)</td>
<td>1.00 (sd=0.00)</td>
<td>0.87 (sd=0.16)</td>
</tr>
<tr>
<td>2-1</td>
<td>1.25 (sd=0.45)</td>
<td>1.17 (sd=0.45)</td>
<td>0.75 (sd=0.35)</td>
<td>0.84 (sd=0.18)</td>
</tr>
<tr>
<td>2-2</td>
<td>1.00 (sd=0.00)</td>
<td>1.24 (sd=0.54)</td>
<td>1.00 (sd=0.00)</td>
<td>0.82 (sd=0.18)</td>
</tr>
<tr>
<td>3-1</td>
<td>1.25 (sd=0.45)</td>
<td>1.35 (sd=0.50)</td>
<td>0.75 (sd=0.35)</td>
<td>0.75 (sd=0.27)</td>
</tr>
<tr>
<td>3-2</td>
<td>2.32 (sd=2.34)</td>
<td>4.77 (sd=9.44)</td>
<td>0.59 (sd=0.06)</td>
<td>0.56 (sd=0.27)</td>
</tr>
<tr>
<td>4</td>
<td>1.00 (sd=0.00)</td>
<td>1.54 (sd=1.20)</td>
<td>1.00 (sd=0.00)</td>
<td>0.81 (sd=0.24)</td>
</tr>
<tr>
<td>5</td>
<td>1.17 (sd=0.39)</td>
<td>1.23 (sd=0.42)</td>
<td>0.83 (sd=0.00)</td>
<td>0.81 (sd=0.18)</td>
</tr>
<tr>
<td>6</td>
<td>1.31 (sd=1.01)</td>
<td>1.31 (sd=1.01)</td>
<td>0.88 (sd=0.00)</td>
<td>0.88 (sd=0.00)</td>
</tr>
</tbody>
</table>
Finally, there were two students who completed all the levels. Table 3 compares how the two students performed in relation to the rest of the group. Those two students have shown higher problem solving skills compared to the group. They also performed well in Level 6 itself, solving most of the exercises on their first try. Level 6 involves looking at the overall story in the form of a picture and connecting it with the compounding 3 numbers. As we discussed before, students who perform this well should not have trouble learning to connect a story to an arithmetic expression. The two students’ good performance on both Level 3 and on Level 6 fits with our model. In other words, it fits with the idea that being able to divide a story into its 3 quantity concepts is key to meaningfully connect a story to an arithmetic expression. However, little can be said with only two subjects. Further experimenting should shed light into this matter.

6. Conclusion

The use of the application by the kindergarten kids has shown to be satisfactory, both in their affective response and in their performance while solving the problems, which was more efficient than gaming the system. This result suggests that:
1. The application is successful at allowing for meaningful interaction with the Triplet Structure Model without the use of text and arithmetic expressions;
2. The students can meaningfully interact with the Triplet Structure Model through the application.

Furthermore, this suggests that the application design for story problems that relies only on images and sound has been effective. It also shows modest signs that students can interact with the structure in word problems even before starting explicitly studying them in primary school grades.

Unfortunately, students had limited time to interact with the application. This resulted in not enough data being available for the later levels. As we gather more data we will have more compelling evidence that kindergarten students can meaningfully interact with the Triplet Structure Model. Lastly, the two students that have completed Level 6 are likely to have a good grasp of the conceptual quantities in the pictures. These students have also shown remarkable performance while assembling the pictures in Level 3. This shows promising signs of a correlation between the ability to understand the conceptual quantities and the ability to understand the story when divided into 3 parts, which would be suggestive of the effectiveness of the Triplet Structure Model.

References


Framework for Building a Thinking Processes Analysis Support System: A Case Study of Belief Conflict Thinking Processes

Yuki HAYASHI*, Kazuhisa SETAa & Mitsuru IKEDAb

a Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan
b School of Knowledge Science, Japan Advanced Institute of Science and Technology, Japan
*hayashi@kis.osakafu-u.ac.jp

Abstract: While the metacognitive skills are the essence of human existence, it is difficult to train thinking skills because they essentially involve cognitive activities and implicit behaviors. This study proposes a framework as a methodology to approach the implicit and chaotic metacognitive processes. The developed framework provides a guidance to develop a system that interprets the users’ thinking processes. As a case study, we develop an application based on the framework to capture the users’ gazing behaviors and actions to modify externalized thoughts in the thinking processes about belief conflict. In addition, we propose an analysis support system for the thinking processes that allows analysts to develop a set of interpretation rules to grasp a part of the users’ metacognitive monitoring and control processes. The study findings indicate the efficiency of the proposed gaze-thinking framework. This is shown through the provided examples of the set interpretation rules and the results of interpreted thinking processes.

Keywords: Metacognitive thinking processes, gaze behaviors, interpretation of thinking processes, thinking of dissolution of belief conflicts

1. Introduction

Metacognition or “thinking about thinking” is the process that corresponds to higher order thinking which involves active control over the cognitive processes. Such metacognitive skills play critical role in successful learning (Livingston, 1997). The ability to clearly express the inherent thoughts in a logical manner has received increasing attention as an essential skill in our social life (Griffin et al., 2012). However, it is challenging to train these thinking skills because they essentially involve cognitive activities and implicit behaviors that are unobservable by others. On the other hand, the proper teaching materials/methodologies for cultivating such skills have not been established yet.

Thinking aloud is a well-known form of an implicit metacognitive thinking process that requires participants to think aloud while solving a problem or performing a task (Jaspers et al, 2004). This method is traditionally applied mainly in psychological and educational research on analyzing cognitive processes. However, some of its limitations have been recognized by researchers in the field; for example, only the information that is actively processed in the working memory can be verbalized (Jääskeläinen, 2010). Additionally, the high cognitive load hinders verbalization by exhausting the available cognitive resources (Miyazaki & Miyazaki, 2004). The thinking-aloud task would also be challenging for the participants with poor self-reflection skills, as it tends to interrupt their natural thinking processes. Stimulated recall is another research method that uses video and audio recordings while the participant is performing a task. The recorded mediums are later reviewed and used as a prompt to get the participants reflections (Fox-Turnbull, 2009). However, this method involves asynchronous work that requires the participants (thinkers) to perform introspective work after they complete their tasks and they might add extra reasons/opinions afterwards.

In order to tackle these problems, many studies that utilize eye-tracking data (gaze data) have been conducted based on psychological perspectives. e.g., verbalization processes (Guan et al., 2006) and normal vs. mindless reading (Reichle et al., 2010). Furthermore, several studies in the research field of intelligent tutoring systems have focused on the gaze features as means to support learning. In general,
they apply supervised machine learning techniques to analyze and predict not only the participants’ learning level (Bondareva et al., 2013) but also their internal affective states such as boredom vs. curiosity (Jaques et al., 2014), and occurrences of confusion (Lallé et al., 2016). However, to the best of our knowledge, there is no established methodology to capture “metacognitive thinking processes” based on the gaze behavior using the eye-movement.

In order to create the common understanding and steadily accumulate the findings of the invisible, shapeless, and complex structure of the thinking processes, the researchers essentially need to reach common ground on “how to capture what types of thinking processes from what kind of gaze behaviors (eye-movement patterns)”. Certainly, it is impossible to use gaze behavior to capture all metacognitive thinking processes in our daily life. This research addresses the difficult but crucial challenge and proposes one of the powerful methodologies to test the hypothesis that a part of the meta-level thinking processes could be captured from the “gaze behaviors” and actions to modify externalized thoughts (“thinking control actions”). The study is based on the assumption that an application allows a user to externalize his or her base-level thinking into a designed interface component.

In this study, we propose a novel framework to develop a system that interprets users’ thinking processes according to their gaze behaviors and thinking control actions. Based on the framework, we introduce a thinking processes analysis support system called thinking processes analyzer which focuses on the thinking processes in dissolution of belief conflicts as a case study. In addition, we demonstrate the interpretation results of data gathered in the context of the knowledge building method workshop for nurses. In Section 2 we discuss the gaze-thinking interpretation framework, and then in Section 3, we explain the details of interpretation rules. Based on the framework, we introduce an implementation example of a thinking processes analyzer in Section 4, and discuss the interpretation results of thinking processes in Section 5.

2. Gaze-Thinking Interpretation Framework

Figure 1 shows a framework for capturing thinking processes from gaze behaviors and thinking control actions originally proposed by Hayashi et al. (2016a). It is known that the thinking processes cannot be observed from the external world. The thinking processes include base-level thinking (Fig. 1(i)) and meta-level thinking (Fig. 1(ii)), where the meta-level thinking monitors (metacognitive monitoring) and controls (metacognitive control) the base-level thinking processes (Nelson, 1990; Hacker et al., 2009). The internal self-conversation about the dissolution of belief conflicts (see Section 4.1) involves thinking about fact, conflict, hypothesis, decision, result, etc., each of which is related to belief conflict that corresponds to base-level thinking activities; for example, “Eri was bullied because she changed her attitude when she interacted with boys” (fact), and “I was bullied because I got along well with Eri” (result). Meta-level thinking targets the base-level thinking; for instance, “comparing certain fact with the reason” and “proving a hypothesis,” are components of metacognitive monitoring while, “modifying the conflict” is an aspect of metacognitive control.

Gaze behaviors are observable activities (Fig. 1(iii)) that can be captured by eye-tracker devices. In reading, the eye moves continuously along a target, running through short rapid movements (saccades) and short stops (fixations) (Barlow, 1952). Meanwhile we need to carefully note the duration times of fixations, we can track gaze behavior such as “gazing at certain object at #t1” and “changing gaze target object at #t3” based on the series of saccades and targets of fixations.

Within this context, we introduce the concept of representation objects, which allow thinkers to externalize their output of base-level thinking activity (representation objects for externalization (Fig. 1(iv)). The objects include areas, text-boxes, buttons, and select-boxes, at the software application level. These objects are observable; hence, we can capture the processes of eye movements and thinking control actions to the base-level thinking representation objects.

In this research, we focus on the isomorphism between the cycle of metacognitive monitoring and control in meta- and base-level thinking in one’s head on one side and the cycle of gaze behaviors and thinking control actions to the representation objects on the other side. Hence, if an application allows a user to externalize his or her output of base-level thinking activity onto the appropriate representation objects (Fig. 1(v)), gaze behaviors and thinking control actions toward the objects indicate a portion of his or her meta-level thinking. As the externalized base-level thought is a portion of base-level thinking
and may be added to or modified through the externalization of internal self-conversation processes, we regard the latter cycle as semi-metacognitive monitoring and control.

In meta-level thinking processes, the monitoring and control of base-level thinking differ depending on the thinking task. For example, when one thinks of the case externalized by another (e.g., correction strategies), he/she needs to recognize, understand, and modify the other’s case. In order to deal with the aspect on the semi-metacognitive level, we introduce “interpretation rules” (Fig. 1(vi)) that express the possible base-/meta-level thinking by gaze behaviors and/or thinking control actions according to the thinking strategies. However, as several meanings can be interpreted from some gaze behaviors representing not only meta-level but also base-level thinking, we consider that the interpretation rules do not determine a unique interpretation for each gaze behavior and thinking control actions.

The interpretation rules in the proposed framework are set by analysts using a thinking processes analysis support system. Thus, we assume a portion of the thinking processes in the metacognitive monitoring and control can be captured by developing a thinking externalizing application with an interface that consists of representation objects each of which is indexed by the base-level thinking of the target task, and by setting interpretation rules for the captured gazing behaviors and thinking control actions of thinkers.

3. Representation Format of Interpretation Rules

Interpretation rules play an important role in the gaze-thinking interpretation framework discussed in the previous section, as they tend to highlight the aspects of the implicit metacognitive thinking processes. The explicitly defined interpretation rules allowed the analysts to understand, share, and compare the interpretation results on a common ground. In this section, we explain the representation format of the interpretation rules. Figure 2 shows a schematic that illustrates the application of interpretation rules to raw data, which includes both the thinker’s gazing behavioral data to representation objects and thinking control action data measured by a thinking externalization application. This process is roughly divided into the following two steps:

**Preprocessing step:** This process unites the measured raw data with noise rejection. First, very short time gazing data are eliminated based on the defined fixation time. Then, similar raw data are checked to determine whether the adjoining data can be unified as gazing behaviors and thinking control actions or not based on the defined time interval.

**Thinking-task-structure-dependent step:** This step applies the interpretation rules to each data interval unified in the previous step. In order to deduce the metacognitive-level thinking processes and obtain various interpretations, we employ forward-processing interpretation rules. In this study, we define two
types of interpretation rules: low-level interpretation rules, which provide primitive interpretations, and high-level interpretation rules, which infer the metacognitive-thinking/control based on the low-level interpretations.

3.1 Low-level Interpretation Rules

In order to provide primitive interpretation $PI_L$ of the measured row data, the low-level interpretation rule $lr_x$ is represented by the following expression 1:

$$lr_x(Act(tr_o_i), PI_L)$$  \hspace{1cm} (1)

where $Act(tr_o_i)$ represents gazing behaviors and thinking control actions processed in the preprocessing step and this function provides certain low-level interpretation of the corresponding interval data. Each of the actions take a certain thinking representation object $tro_i$ as an argument. $GazeAt(trro_i)$ is prepared for gazing behaviors, and $Keypress(trro_i)$, $Delete(trro_i)$, $Press(trro_i)$, etc. are used as the elements of thinking control actions.

3.2 High-level Interpretation Rules

The high-level interpretation rules raise the low-level interpretations to the metacognitive monitoring/control levels. A high-level interpretation rule is represented by the following expression 2:

$$hr_x(Cond, HI_L)$$  \hspace{1cm} (2)

where $hr_x$ represents the function that detects the matching intervals of the condition $Cond$, and gives a high-level interpretation $HI_L$ of the intervals. $Cond$ consists of one or more sets of conditional function(s). Table 1 summarizes the conditional functions and their explanations that are used to set $Cond$ based on the possible relations between two intervals (Allen, 1983) to be utilized for time series analysis. As shown in Table 1, arguments $L_1$ and $L_2$ represent the results of data intervals deduced by the low-/high-level interpretation rules. $All(L_1)$ represents all data of $L_1$, $BEFORE(L_1, L_2, t)$ indicates the interval between the starting time of $L_1$ and finishing time of $L_2$ if the interval of $L_1$ is before that of $L_2$ within $t$ ms. $OVERLAPS(L_1, L_2)$ detects the overlapped intervals between $L_1$ and $L_2$, and $DURING(L_1, L_2)$ extracts the intervals $L_1$ if $L_2$ is also observed simultaneously.

Note that the appropriateness of the detected interpretations depends on the analysts, since the thinking itself cannot be observed from the external world. Hence, we cannot obtain the genuinely correct answers of one’s base-/meta-level thinking processes. Within this context, the analysts can explicitly set their interpretation rules that express what sort of metacognitive activities can be captured by the types of gazing behaviors.

Table 1: Conditional Functions for Time Series Analysis.

<table>
<thead>
<tr>
<th>Function</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$All(L_1)$</td>
<td>all $L_1$</td>
</tr>
<tr>
<td>$BEFORE(L_1, L_2, t)$</td>
<td>$L_1$ takes place before $L_2$ within $t$ ms.</td>
</tr>
<tr>
<td>$OVERLAPS(L_1, L_2)$</td>
<td>$L_1$ overlaps with $L_2$</td>
</tr>
<tr>
<td>$DURING(L_1, L_2)$</td>
<td>$L_1$ during $L_2$</td>
</tr>
</tbody>
</table>
toward the kind of base-level thinking representation objects and thinking control actions. Accurate examples of applying low-level and high-level interpretation rules are described in Section 5.

4. Example of a Thinking Processes Analysis Support System

In this section, we explain the analysis support system developed based on the gaze-thinking interpretation framework. First, in Section 4.1, we clarify the target thinking processes and summarize the thinking training environment developed by Chen et al. (2011) that forms the basis of our system. In Section 4.2, we introduce an application that can capture the sequence of the user’s gazing behavioral data to representation objects and thinking control actions during his/her meta-level thinking processes. Then, in Section 4.3, we introduce an analysis support system of thinking processes which allows analysts to set their interpretation rules as discussed in Section 3 and confirm the results of metacognitive monitoring/control levels’ thinking processes through the visualized timeline window.

4.1 Target Thinking Processes

Ito (2009) reviewed the effect of the thought verbalization as a learning strategy and proposed a model on verbalization to meet the learning goals. The model describes the sequence of three phases: description (cyclic state of verbalizing one’s thought based on personal experiences), cognitive-conflict (state of facing the conflict through the verbalization of one’s thought and interaction with others), and knowledge-building (cyclic state to resolve the conflict states). Throughout these phases, if the learners actively and logically think about the problem they face, their thinking processes become clear and they tend to develop more sophisticated thinking skills.

Based on the verbalization model, Chen et al. (2011) proposed a thinking training environment called Sizhi to improve the learners’ thinking about thinking skills. Sizhi targets the thinking processes about belief conflict, which is defined as a fundamental confrontation caused by the situations individuals face when their beliefs are questioned (Kyougoku, 2011; Kyougoku et al., 2015). The interface is designed to logically verbalize one’s own thinking and that of others by reflecting on one’s own thinking process in a logical manner, in order to find meaningful conflicts.

Sizhi has been continuously used in previous research to conduct educational programs for first-year bachelor students many of whom are beginners to meta-thinking (Seta et al., 2013), in addition to hospital nurses who suffer from belief conflict in the medical field (Kanou et al., 2013). The nursing education program, not only involved nurses externalize their own belief conflict, but also offered skill training (by an expert on meta-thinking) who corrects and reviews the outcome of the cases externalized by nurses. The reported results revealed that the Sizhi can effectively cultivate the metacognitive skills of the learners.

On the other hand, as the quality of the externalized/corrected thought (i.e. collective statements and their logic) varies depending on the learners/correctors, the thinking processes leading to externalization of thought is still implicit. If we could grasp a part of one’s implicit thinking processes in metacognitive monitoring and control level, we would further advance the learning analytics of certain thinking tasks, e.g., the difference between the thinking externalization/correction processes of learners/correctors and whether they are deeply finding meaningful conflicts or not. In addition, if the processes can be modeled in human-understandable levels, intellectual support can be provided to foster the thinking about thinking skills.

4.2 Eye-Sizhi

Based on the proposed framework, we have proposed an Eye-Sizhi application, as a case study to capture the user’s gaze behaviors and thinking control actions toward representation objects of base-level thinking (Hayashi et al., 2016b). Figure 3 represents the interface of the Eye-Sizhi, which follows the design principle of Sizhi. The application is used as instructional equipment that is premised to be utilized in the context of thinking training. The interface consists of four thinking areas: ‘A’s-thinking’ that denotes one’s own thinking, ‘B’s-thinking’ denotes opponent’s thinking, ‘conflict’ denotes the difference between A’s- and B’s-thinking, and ‘knowledge-building’ denotes dissolving the belief conflict.
In the application, the user can add/delete their statements using the statement edit buttons and input the statement text into the textbox by specifically selecting the pre-defined Sizhi-tags such as *fact*, *hypothesize*, *decision*, *assumption*, and *policy* to logically verbalize his or her output of base-level thinking activity. The user can also associate other statements as 'references' to express the basis for adding the statement. In order to help the learner gain deep insight into the belief conflicts in the conflict areas, the user is allowed to select only one policy statement from each of the statements externalized in the A’s- and B’s-thinking areas, and express the belief conflict in the conflict textbox.

In order to capture the users’ gazing target data, Eye-Sizhi is implemented using a screen-based eye-tracker, which distinguishes and records the thinking area in the interface at which a user is looking by setting the area of interest (AOI) regions of the representation objects. The application recognizes the four thinking areas (*A’s-thinking*, *B’s-thinking*, *conflict*, and *knowledge-building*), each statement area itself and the included components (Sizhi-tag combo-box, reference label, and textbox), the conflict textbox area, and the edit buttons. The application records the user’s activity on millisecond time scale, which includes the user’s gaze events and thinking control actions (i.e. keyboard and mouse click events).

### 4.3 Thinking Processes Analyzer

We developed a system called thinking processes analyzer which accepts the Eye-Sizhi records as an input file and allows the analysts to set their low-/high-level interpretation rules. Figure 4 illustrates the interface of the analyzer. The interface mainly consists of the following three areas and the timeline visualization window for the results of set interpretations.

> **(1) Low-level interpretation processing area:** In this area, the analysts can set three types of time intervals on the millisecond time scale: (i) the duration times of fixations, (ii) time interval for unifying the similar adjoining data as gazing behaviors and externalization actions (gazing at same thinking/statement areas, typing out actions), (iii) time interval for the instantly recorded actions such as clicking buttons in order to be assessed and visualized in the visualization window. After the settings have been established, the analyzer reads the raw data of imputing the Eye-Sizhi log file and extracts each gazing behavioral and thinking control action data. In addition, it annotates the preset low-level interpretation rules (e.g., "Understanding_Thinking.all" to the intervals of gazing at each thinking area, "Modifying_Statements.all" to the intervals of typing out actions in statements, etc. in case of correction strategy). These processes correspond to the processing of row data into low-level data interpretation, as shown in in Fig. 2.

The results of the low-level interpretation rules are displayed in the visualization window where each of the interpreted intervals is allocated in a timeline style. The timelines can be accessed by dragging the operation, and the analysts can check the detail contents of each interval by hovering a mouse cursor over the interval. The lower right part of Fig. 4 illustrates the situation between section 1:00 to 2:00 during the session (about 42 minutes) that can be viewed by zooming in and hovering the mouse cursor over a certain gazing interval of a thinking statement. In this case, the statement information (ID: 36, Sizhi-tag: 結果 (result), basis: statement whose ID is 35) and the gazing time interval are shown on the tooltip.
This area displays the generalized results of the Sizhi input record file in which the analysts can confirm the total time of how a learner/corrector gazes at each thinking panel/statement; how many times he/she inputs the text, adds/deletes statements, changes Sizhi-tags; etc., throughout the session.

(3) High-level interpretation processing area: This area allows the analysts to set their high-level interpretation rules to the results of Process (1), and analyze the results of set rules obtained from the timeline window. This process corresponds to the cycle of applying high-level interpretation rules shown in Fig. 2.

A high-level interpretation tag-name for the new rule is set in the textbox area in Fig. 4(a). Functions of the condition part and the argument(s), which corresponds to the result of low-level interpretation rules, can be selected from the combo-boxes in Fig. 4(b). In order to provide detailed specifications of the base-level thinking of statements, the analysts can set a particular type of Sizhi-tag in case the argument class is a statement. If necessary, they also specify the type of Sizhi-tag of the basis statements and statement ID as optional settings for Sizhi. The area shown in Fig. 4(c) allows the analysts to check and add the set interpretation rules. The added interpretation rules are listed in the rule list table (Fig. 4(d)), which includes the low-level interpretation rules obtained in Process (1). The tag-name of each interpretation rule appears in combo-box items, as shown in Fig. 4(b). After setting a series of high-level interpretation rules, the corresponding intervals of the rules are extracted in the area shown in Fig. 4(e) that can be visualized in the timeline visualization window (Fig. 6) apart from the window obtained in Process (1).

As explained above, the raw data about belief conflicts are manipulated by the developed thinking processes analyzer using Processes (1) to (3). This allows the analysts to develop their interpretation rules regarding the raw data measured during the thinking processes. They can visually capture the results through the visualization window and add further high-level interpretation rules in order to grasp a part of the learners/correctors metacognitive monitoring and control processes.
5. Example of Interpreted Thinking Processes

5.1 Data Collection

In order to examine the efficiency of the proposed gaze-thinking framework in interpreting the metacognitive thinking processes, we collected the correction processes data using Eye-Sizhi described in Section 4.2. In the correction situation, the correctors are expected start with understanding the externalized thought and they mainly focus on the metacognitive activities under the objective of the correction. Thus, we expect a part of the metacognitive monitoring and control processes is reflected in gaze behaviors and thinking control actions that take the form of typing out actions as semi-metacognitive monitoring and control.

The knowledge building method workshop for nurses (Kanou et al., 2013) is selected as a case study. Two correctors, who have some previous experience correcting hospital nurses’ cases, were asked to correct the cases of the nurses about their experiences of belief conflict in the medical field using Eye-Sizhi. Eye-Sizhi is continuously used in the knowledge building workshop since 2015. Currently, 26 correction processes datasets exist performed by two correctors for 14 nurses’ cases. The correction task was continued until the correctors were fully satisfied. The guiding principle of the correction was to check whether the externalized A’s- and B’s-thinking are logically correct and reveal meaningful belief conflicts.

5.2 Results and Discussion

Figure 5 illustrates some examples of visualization windows showing the applied low-level interpretation rules to recorded correction processes of three cases. The applied interpretation rules are listed in Table 2. The timeline shows the intervals of ①: gazing at types of thinking panels (lr1), ②: gazing at the contents in conflict area (lr2: areas of policy statement of A’s-/B’s-thinking and conflict textbox), ③: gazing at statements (lr3), ④: typing out statement actions (lr4), as shown in Fig. 5. The visualized results of ① and ② indicate that the sequences of gazing are not always chaotic; instead, they have a certain time duration to understand and correct the nurses’ externalized thought. This explains how to interpret the succession of correctors’ monitoring and control

Table 2: Examples of Low-level Interpretation Rules.

<table>
<thead>
<tr>
<th>Low-level Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>lr1 (GazeAt(Thinking_area[all]), &quot;Understanding_Thinking.all&quot;)</td>
</tr>
<tr>
<td>lr2 (GazeAt(Conflict_area[all]), &quot;Understanding_Conflict.all&quot;)</td>
</tr>
<tr>
<td>lr3 (GazeAt(Statements[all]), &quot;Understanding_Statements.all&quot;)</td>
</tr>
<tr>
<td>lr4 (Keypress(Statements[all]), &quot;Modifying_Statements.all&quot;)</td>
</tr>
</tbody>
</table>
processes. In addition, we also found these three cases started by focusing on understanding the contents in conflict area (purple colored intervals).

Figure 6 shows examples of adapting high-level interpretation rules. In this case, a set of rules listed in Table 3 is applied to Case I shown in Fig. 5. Figure 6(i) is the timeline that corresponds to the intervals in which the corrector was comparing A’s-thinking with B’s-thinking by using high-level interpretation rule $hr_1$, whose condition includes low-level interpretation rule $lr_1$. Based on this timeline, we can grasp the corrector’s metacognitive monitoring processes of comparing one’s thinking with that of another in the first half and last of the session (the blue dotted-line area in Fig. 6(ii)). The timelines visualized in Fig. 6(ii) and (iii) are derived from high-level interpretation rules $hr_2$ and $hr_3$. These rules aim to clarify the corrector’s understanding processes intervals ($hr_2$) and modify the actions ($hr_3$) related to the policy statements that form the root of belief conflict. The results of these rules set the high-level interpretation rule $hr_4$ to capture the intervals of metacognitive thinking processes. These intervals involve corrector deliberation about the policy statement correction, as the policy statement intervals are understood before the statements are modified (the green dotted-line area in Fig. 6(iv)).

The above discussion demonstrates the results of applying high-level interpretation rules to uncover a portion of the thinking processes on the metacognitive thinking level in the context of the thinking task about belief conflict. However, it is difficult to determine whether the result of interpretations is truly correct. Hence, in order to steadily approach the intangible human thinking processes, we believe it is necessary to share the knowledge about what forms of meta-level thinking processes are to be captured by what types of gaze behaviors toward what kind of base-level thinking representation objects on the basis of our proposed framework.

6. Conclusion

In this study, we proposed a gaze-thinking interpretation framework as a promising methodology to capture a part of the learners’ metacognitive thinking processes. The fundamental idea of the framework is based on the isomorphism between the cycle of metacognitive monitoring and control in the thinking processes in one’s head and the cycle of gaze behaviors and thinking control actions to the base-level thinking representation objects.

As a practical application of the framework, we developed a thinking processes analyzer that allows analysts to analyze the belief conflict thinking processes by setting low-/high-level interpretation rules. To examine the efficiency of the proposed gaze-thinking framework, we demonstrated the results of applying interpretation rules using limited study sample. The future research will provide a more detailed analysis of the data collected from the knowledge building method workshop for nurses to clarify the difference between the thinking correction processes adopted by correctors based on the developed analyzer.
Acknowledgements

This work is supported by The Telecommunications Advancement Foundation (TAF), JSPS KAKENHI Grant Numbers JP26870588 and JP17K18017. The authors also would like to thank N. Matsuda for the technical assistance during data collection.

References


Usability and Learning Effect Evaluations of an Electrical Note-Taking Support System with Speech Processing Technologies

Hiromitsu NISHIZAKI\textsuperscript{a} & Yosuke NARITA\textsuperscript{b}

\textsuperscript{a} Integrated Graduate School of Research, University of Yamanashi, Japan
\textsuperscript{b} Integrated Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi, Japan
*hnishi@yamanashi.ac.jp

Abstract: This paper describes the evaluation of an electronic note-taking support system that works on a tablet/laptop computer in a real classroom environment. The distinguishing characteristic of the proposed electronic note-taking support system is the incorporation of a few state-of-the-art speech processing technologies. One is deep-neural-network-based automatic speech recognition for transcribing talks. Such transcriptions are very useful for not only note-taking but also recovering missing words. The other speech processing technology is keyword search in the recorded speech. We had already developed the system previously and evaluated it. In this study, we improved the previous support system from the viewpoints of usability and visibility of the system user interface. The improved system was evaluated in the subjective experiment in which listeners (students) who attended the real lectures. In the experiment, not only the usability of the system but also the learning effect for knowledge acquisition were evaluated. The experimental results showed that the improved note-taking support system improved usability and visibility compared to the previous system and the traditional note-taking method. Moreover, the students claimed that the system helped them acquire knowledge about the lectures they attended.

Keywords: Learning effect, note-taking support system, user interface, usability evaluation

1. Introduction

In our daily life, we have many opportunities to take notes such as in classroom lectures and meetings. When we take notes, we usually use a paper notebook with a pen (or pencil) or an electronic device such as a tablet or a laptop personal computer (PC). However, often, we miss words (contents) uttered by a speaker.

Nowadays, most lecturers make slides for lectures using presentation software such as Microsoft PowerPoint\textsuperscript{®}. Comparing a lecture with presentation software to a lecture with a blackboard, most lecturers using presentation software talk and turn slides faster than those not using presentation software. Therefore, most listeners (lecture attendees) may at times miss parts of a talk by a lecturer using presentation software. Moreover, some listeners squiggle characters on a paper notebook. These listeners may not be able to understand the characters, some of which could be messy.

Various types of note-taking support systems and applications (Zyto, et al., 2012), (Livescribe, Inc., 2017), (Google Inc., 2017), (Echo360, 2017) have been developed for solving these problems. They provide listening support and can be used to take notes efficiently.

For example, “NB” (Zyto, et al., 2012) provides a sharing function among listeners who take the same class. The focus of this application is on the sharing function; therefore, it is unsuitable for taking notes at a lecture in real-time. “Livescribe 3 SmartPens” (Livescribe, Inc., 2017) is a commercial note-taking tool that consists of a special pen and a special notebook. This tool records an entire talk by a lecturer; therefore, the user can listen to the talk when reviewing the notes taken using the tool. The “Google Keep” (Google Inc., 2017) application allows users to take notes on portable devices. Users can easily insert a photo in a note and superimpose text onto it. In addition, voices can be recorded and...
stored in the application. Finally, “Echo360” is an enterprise system for classroom lectures that can record the video and the speech of a lecturer and distribute the recorded content to students. Students can review the lecture video and communicate with the lecturer through the system whenever they like. However, Echo360 does not have any note-taking function.

Although there are a few systems and applications that allow users to review lectures and perform and note-taking tasks, none of these applications/systems have the following characteristics from the viewpoint of classroom use:

- Ease of note-taking
- High visibility of note contents
- Recovering missing words

Therefore, Ota, Nishizaki, and Sekiguchi (2011) developed a note-taking support system for recording classroom lectures. Listeners (note takers) can efficiently use captured images of slides projected onto a screen or figures/characters written on a blackboard for note-taking. In addition, they can also use transcriptions, which are generated by an automatic speech recognition (ASR) system, for note-taking. These characteristics the support system enable lecture listeners to make visible note content without missing any of the lecturer's talk because they can concentrate on the lecture. The system has a function that shows the ASR result on a graphical user interface. Therefore, listeners can not only focus on the lecturer's talk but also watch “visible words” by the lecturer. In addition, the visible words can be used to create a memo by touching the displayed words.

The support system has other useful functions to support note-takers. For example, a spoken term detection (STD) technology (Akiba et al., 2011), a spoken keyword search/extraction technique, can find the occurrence positions of a query term (keyword) in a recorded lecture talk. This STD enables us to cue a recorded talk easily when a listener wants to review the note content. The efficiency of each useful function implemented in the support system has been already evaluated in previous works (Ota, Nishizaki and Sekiguchi, 2011), (Yonekura et al., 2013).

However, the previous evaluation results showed that the support system was not very handle-able, in other words, it is hard for users to operate the system, and the user interface of the system was not very good compared to the traditional note-taking method of using a pen and paper.

Our note-taking system provides four useful functions, namely, capturing images of projected slides or hand-written characters on a blackboard, ASR of a lecturer’s speech, hand-written and one-touch memos using visible words generated by ASR, and keyword search. These functions are unique to our note-taking system. Moreover, no studies have investigated the effects of note-taking support systems and applications such as the one described above on learning. Therefore, in this study, we experimentally demonstrate the effectiveness of learning using such applications by asking a few students to use our note-taking system in a classroom.

We improve the user-interface of the previous note-taking support system (Yonekura et al., 2013) and re-evaluate the improved system in a real classroom (the previous system was not evaluated in a real classroom but in a simulated classroom). In addition, we investigate the effect of the proposed note-taking system on learning.

The evaluation results obtained in the real environment (real classroom) showed that the latest support system improved usability and visibility compared to the previous system and the traditional note-taking method. Moreover, the evaluation results show that our system helped users acquire more knowledge about the lecture compared to a traditional note-taking method.

2. Note-taking Support System

2.1 System Outline

Figures 1 and 2 show the processing flow of our note-taking support system and a screenshot of the note-taking device, respectively. The newest version of the system, which has been upgraded from previously reported versions (Ota, Nishizaki and Sekiguchi, 2011), (Yonekura et al., 2013) by following the result of previous subjective experiments, was used in this study. As shown in Figure 1, a
talk by a lecturer is recorded automatically by a server, which also captures screenshots of the projector screen (or a blackboard) at intervals of 3 s. After the speech is stored on the server, it is transcribed by an
ASR system. The transcription is then transferred to a note-taking device. At this time, the speech is also converted to a PTN-formed index (Natori et al., 2013) for STD by using the ten ASR systems.

The word sequences generated by the ASR system are filtered by rules on the note-taking device and are displayed on the note-taking device. A user can record a note by simply touching or tracing words on the screen. Therefore, handwriting and keyboard operations are unnecessary. The entire speech is recorded on the server, and the locations of recognized words in the speech are identified using word-to-speech alignment information generated by ASR. After note-taking is completed, the recorded speech is transferred to the device. Therefore, users can easily play the speech back from any chosen point. It is not necessary to listen to the entire speech when checking a note.

2.2 Speech Interface

Our note-taking support system has a speech interface as well. This speech interface captures speech, records notes, and performs speech recognition. The recognized words are displayed on the screen. Users can record a note by simply touching or tracing relevant words with their finger. Therefore, this reduces the burden on the user of recording notes while listening to a speech.

However, this system depends on speech recognition. If the words the user wants to associate a note with are not recognized correctly, the note cannot be recorded. Speech recognition errors are unavoidable. Therefore, it is undesirable to completely trust the ASR technology.

The key concept of the system is to use ASR as an accessory function. The ASR system is designed to help users record notes; however, it must disturb their work as little as possible. If the words that the user wants to associate a note with are not recognized correctly, the user does not need to touch or trace the words on the screen. Rather, the user can record a note using a keyboard or an electronic pen. This concept of using ASR as an accessory function differs from other systems with speech interfaces. The usability of those system depends entirely on the output of the ASR system, and the system usability worsens if ASR performs poorly.

Speech is recorded simultaneously as ASR operates. Therefore, users can listen to the speech many times if needed. Each annotated word has information that describes where the word is located in the speech. Therefore, system users can easily play the speech starting from any specified word.

2.3 Keyboard and Handwritten Input

As described in Section 2.2, a note-taker can record a note by using a (hardware or software) keyboard or an electronic pen in addition to ASR. If words are transcribed faultily by ASR, the user may input words by using a keyboard or by writing.

A note-taker can draw graphics, such as underlines, on the screen. They can also write words and circle finger-touched or handwritten words for emphasis.

In this study, all handwritten and keyboard-input words are called objects. If an object is drawn on the system screen, it is correlated with the relative time from the beginning of the recorded speech. Therefore, if a user touches an object, they can listen to the speech starting from the object-specified time.

2.4 Recorded Note Review

A user can see a recorded note while listening to the recorded speech. As mentioned in Sections 2.2 and 2.3, a system user can also play speech starting from the time specified by the location of the word or object.

In addition to the notes recorded by the user, all recognized words from the speech are stored in the system. This enables the user to search for a specified word from recorded speech. When the user enters a search term in the search window, the locations of the word are identified if it is recognized correctly.

However, users cannot search for words transcribed incorrectly. To avoid this, we have incorporated the STD technique (Natori, et al., 2013), which is sufficiently robust to speech recognition errors.
2.5 Keyword Search Interface

When a user searches a queried term using the STD engine, the search results are shown on the speech playback seek bar. Figure 3 shows an example of the interface for outputting STD search results. The search results are shown in list form. Furthermore, the expected locations of the queried term, as determined by the STD engine, are indicated on the seek bar (small circle).

2.6 Other functions

Our note-taking system has the following other useful functions:

- Overhead-view of all pages
  A user can display a thumbnail list of all pages. Therefore, they can smoothly move to the target page when they want to review a note.
- Marking important pages (bookmark)
  A user can bookmark favorite or important pages in a note.
- Register and emphasize important words from an ASR transcription
  A user can register important (or key) words into the system in advance. If a lecturer utters these words, the system alerts the user and the words are emphasized with bold font and displayed in the ASR result area.
- Change color of pen easily by touching a “<” or “>” button (see Figure 2). The color of the pen is set to blue by default, and the only other registered color is red. Of course, a user can register other colors freely. It is well-known that blue and red colors are brain-stimulating; therefore, these colors enhance the effectiveness of learning (Mehta and Zhu 2009).

3. Improvement over Previous Note-taking System

We evaluated the previous note-taking system (Yonekura et al., 2013). The evaluation results showed that the previous system reduced users’ note-taking burden compared to the traditional (hand-written) method of note-taking. However, it did not fare well in terms of operability and the visibility. Therefore, we had to improve the previous system. A few test subjects indicated the following points of improvement:

- Expansion of note-taking space
- Show grid lines
- Ability to change color of characters easily
Create a new page when a user captures an image

Improve ASR performance

We have refined the previous note-taking system based on the above suggestions, and we will compare the usability of the revised version of the system to the previous one.

4. System Evaluation

We performed a subjective experiment to evaluate the usability of the newest note-taking support system and the visibility of the user interface on the system in a real classroom environment.

In the experiment, we test the usability of the system and its effect on learning in lectures by comparing three note-taking methods: the newest support system, the previous system, and the traditional note-taking method (pen and paper).

![System evaluation design in subjective experiment.](image)

Figure 4. System evaluation design in subjective experiment.

Table 1: Experimental conditions for note-taking evaluation.

<table>
<thead>
<tr>
<th>Lecture name</th>
<th>Fundamentals of machine learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lectures</td>
<td>Five (only one lecture was used for comparison of the newest system and the previous system)</td>
</tr>
<tr>
<td>Duration time</td>
<td>90 minutes for all lectures</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>12 (undergraduate students who majoring in Mechatronics engineering)</td>
</tr>
<tr>
<td>ASR system</td>
<td>Julius rev. 4.3.2 (Lee &amp; Kawahara, 2009)</td>
</tr>
<tr>
<td>Acoustic model</td>
<td>Triphone-based Deep Neural Network (DNN)-Hidden Markov Model (HMM) (Hinton et al., 2016)</td>
</tr>
<tr>
<td>Language model</td>
<td>Word-trigram trained from Corpus of Spontaneous Japanese (Maekawa, 2013)</td>
</tr>
</tbody>
</table>

4.1 Experimental Setup

Figure 4 shows the outline of the subjective experiment. In addition, Table 1 lists the conditions of the note-taking evaluation. The subjects were 12 undergraduate students majoring in Mechatronics...
All students were not accustomed to operating a computer, but they used their smartphones regularly. All students attended the four classes (lectures) on fundamentals of machine learning, and each lecture lasted 90 min. After the lecture (except for the fourth lecture), all students took a mini-examination related to the lecture and answered a questionnaire on system usage. They were allowed to review note contents (generated by the system or hand-written) during the examination. The scores of the mini-examinations were a good indicator of the effect of the system on learning during the lectures. The questionnaire results will be used to assess system usability.

In these four lectures, all subjects used the newest (refined) note-taking support system and the traditional method. For a comparison of the newest and the previous systems, we scheduled an additional lecture in which the subjects evaluated the previous system.

We instructed (for a few minutes) the all students about the use of the note-taking system before the lectures. However, all students did not practice system operation for long. Instead, we provided a manual in which system operating instructions were explained to all students. They were free to read the manual whenever they liked during the lectures.

Table 2 summarizes the number of students using the various note-taking methods. As shown in Table 2, three or four students per lecture (except for the 5th lecture) used the newest system. All students evaluated the newest system once. For assessment of the improvement in the user interface, we compared the newest note-taking system and the previous version. Five students used the previous system in the fifth lecture. All five students had used the newest system beforehand.

<table>
<thead>
<tr>
<th></th>
<th>Newest system</th>
<th>Paper pen</th>
<th>Previous system</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st lecture</td>
<td>4</td>
<td>10</td>
<td>N/A</td>
</tr>
<tr>
<td>2nd lecture</td>
<td>3</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>3rd lecture</td>
<td>4</td>
<td>7</td>
<td>N/A</td>
</tr>
<tr>
<td>4th lecture</td>
<td>4</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>5th lecture</td>
<td>N/A</td>
<td>N/A</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3: Objective evaluation results.

<table>
<thead>
<tr>
<th></th>
<th>Newest system</th>
<th>Paper pen</th>
<th>Previous system</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many times missed lecturer’s utterances [times]</td>
<td>2.9</td>
<td>4.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Time rate of listening lecturer’s talk in concentration [%]</td>
<td>73.1 %</td>
<td>55.7 %</td>
<td>66.6 %</td>
</tr>
<tr>
<td>Ease of reviewing note [10-step evaluation]</td>
<td>9.4</td>
<td>5.0</td>
<td>8.7</td>
</tr>
<tr>
<td>Usability of systems</td>
<td>8.4</td>
<td>N/A</td>
<td>7.0</td>
</tr>
<tr>
<td>Visibility of user interface</td>
<td>8.4</td>
<td>N/A</td>
<td>6.7</td>
</tr>
</tbody>
</table>

4.2 Experimental Results

4.2.1 Objective Evaluation

Table 3 summarizes the objective results of the questionnaires in which we assessed five items as follows:
- How many times did each student miss the lecturer’s utterances (average of all students)?
- Time rate of listening lecturer’s talk in terms of concentration (subjective)
- Ease of reviewing notes
- System usability (user-friendliness)
- Visibility of graphical user interface (how easy was it to catch the ASR-based caption and how easy was it to review recorded notes?)
As can be inferred from Table 3, the newest system bested the paper-based (traditional) note-taking method on all evaluation items. Because electronic systems can record and transcribe the entire lecture by means of the ASR system, the students who used the systems were able to recover missing words. In addition, the students could capture images of the slides projected on the screen easily; therefore, they concentrated on listening to the lecture.

By contrast, in the comparison of the newest system with the previous one, the new system bested the previous one on all evaluation items. This result shows that the previous system was improved successfully. The upgrades in terms of the five points described in Section 3 were very effective for improving the usability and the visibility of the user interface of the note-taking system.

4.2.2 Evaluation on Learning Effect

Table 4 shows the accuracy rates of the mini-examinations conducted soon after the end of each lecture. As can be inferred from Table 4, the students who used the note-taking system obtained better results in each examination. From Table 3, when the students used the electronic note-taking system, the time rate of concentrating on listening to the lecture was increased drastically compared to that with the hand-written method. In other words, the students who used the system could concentrate on the lecture, resulting in improvements in their examination scores. In addition, the ASR-based captions allowed the students to easily create a memo on a captured image. They could write words or phrases in the note-editing space by using ASR-based captions. This let the students focus on the talk. From the results, the electronic note-taking system helped the subjects study the topic of the lecture more effectively compared to the traditional method of taking notes using pen and paper.

Table 5 also shows the accuracy rates of mini-examinations conducted two months after the end of the fourth lecture. All questions were the same as the ones in Table 4. Different from the examination in Table 4, the students were not allowed to review the notes made by themselves. They answered the questions from their memory only. As may well be expected, the accuracy rates in all examinations degraded compared to the results in Table 4. However, the subjects who used the system scored significantly higher than the other subjects in two examinations (1st and 2nd examinations). However, in the 3rd lecture, the average of the examination scores did not change compared to that in Table 4 (from 66.7 to 66.7). This is because the students who did not use the system studied deep learning programming after finishing the lectures, while the questions in the 3rd examination were related to autoencoders and recurrent neural networks. In other words, the subjects gained knowledge about the 3rd examination from elsewhere. This is a special case. The results in Table 5 indicate that our system was very useful for acquiring knowledge about lectures, even when considering the special case.

Table 4: Mini-examination results soon after end of lectures (accuracy rates [%]).

<table>
<thead>
<tr>
<th></th>
<th>w/ System</th>
<th>w/o System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st lecture</td>
<td>65.0</td>
<td>56.4</td>
</tr>
<tr>
<td>2nd lecture</td>
<td>100</td>
<td>71.2</td>
</tr>
<tr>
<td>3rd lecture</td>
<td>95.8</td>
<td>66.7</td>
</tr>
<tr>
<td>Average</td>
<td>88.9</td>
<td>63.7</td>
</tr>
</tbody>
</table>

Table 5: Mini-examination results about two months after end of lectures (accuracy rates [%]).

<table>
<thead>
<tr>
<th></th>
<th>w/ System</th>
<th>w/o System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st lecture</td>
<td>53.3</td>
<td>40.0</td>
</tr>
<tr>
<td>2nd lecture</td>
<td>78.6</td>
<td>28.6</td>
</tr>
<tr>
<td>3rd lecture</td>
<td>61.1</td>
<td>66.7</td>
</tr>
<tr>
<td>Average</td>
<td>64.3</td>
<td>45.1</td>
</tr>
</tbody>
</table>
5. Conclusions

This paper described evaluations of an electronic note-taking support system in a real classroom environment. The support system allows users to take notes quickly by using captured images of slides projected on a screen and transcriptions created by the ASR system. Recording and transcription of a lecturer's voice by the ASR system can help a note-taker to not miss the lecturer's talk. Therefore, a note-taker can concentrate on the lecture to a greater extent than they would be able to do with the traditional note-taking method.

The experimental results of usability and visibility showed that our system provides a better user interface for creating or reviewing notes. Our system did not prevent the system users from listening to the lectures. In addition, the results of the examination scores claimed that the use of our system helped users to create memories of lecture contents, and finally, the system facilitated the students in acquiring knowledge.

In this paper, we evaluated the system on five lectures of one topic (machine learning). In future works, we will evaluate the system in other lectures and we will increase the number of students.

Acknowledgements

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Grant Number 17H01977 and Grant-in-Aid for Scientific Research (C) Grants Number 15K00254.

References

Student Placement Predictor for Programming Class Using Classes Attitude, Psychological Scale, and Code Metrics

Ryosuke ISHIZUÉa*, Kazunori SAKAMOTOb, Hironori WASHIZAKIa & Yoshiaki FUKAZAWA a

aDepartment of Science and Engineering, Waseda University, Japan
bNational Institute of Informatics / JST PRESTO, Japan
*ishizue@ruri.waseda.jp

Abstract: It is often necessary to divide a class according to students’ skill level and motivation to learn. This process is burdensome for teachers because they must prepare, implement, and evaluate a placement examination. This paper tries to predict the placement results via machine learning from some materials without such an examination. The explanatory variables are 1. Psychological Scale, 2. Programming Task, and 3. Student-answered Questionnaire. The participants are university students enrolled in a Java programming class. The target variable is the placement result based on an examination by a teacher of the class. Our classification model with Decision Tree has an F-measure of 0.937. We found that the set of the following explanatory variables can yield the best F-measure (0.937): (1) Class Fan Out Complexity, (2) Practical utility value, (3) Difficulty Level 4 (AOJ), (4) Difficulty Level 3 (AOJ), (5) Interest value, and (6) Never-Give-Up Attitude.

Keywords: machine-learning, programming class, placement, psychological scale

1. Introduction

It is often necessary to divide students into an advanced class and an intermediate class based on skill level, motivation to learn, etc. However, dividing students is burdensome on a teacher because the teacher must prepare, implement, and evaluate the examination (e.g., placement test or questionnaire about requests regarding class level) to assess students’ ability. Moreover, when the teacher conducts such questionnaire, its interpretation depends on the teacher individually. It causes problems in a class where two or more teachers are assigned or when the teacher changes. Additionally, there are several other problems. For example, some students only memorize the answers of past examinations, while other students cram for a test in one night. This paper aims to properly place students using a method easier than the traditional time-consuming examination.

We focus on a class for second-year undergraduate students learning to program in Java at Waseda University. In this class, students are divided into an advanced class and an intermediate class about a month after the semester begins. Students complete a placement examination by the teacher. In this paper, we try to substitute the examination with a questionnaire, which asks students about their class attitude and the result of a programming task in class. This information is then used to create a machine-learning model to predict the placement results. The explanatory variables are 1. Psychological Scale, 2. Programming Task, and 3. Student-answered Questionnaire. The classification model has precision, recall, and an F-measure of 0.937. Additionally, we evaluate the effects of the explanatory variables on the placement results.

The contributions of this paper are:

- We investigate factors affecting the placement results: (1) Class Fan Out Complexity, (2) Practical utility value, (3) Difficulty Level 4 (AOJ), (4) Difficulty Level 3 (AOJ), (5) Interest value, and (6) Never-Give-Up Attitude.
- We create a model with Decision Tree which has an F-measure of 0.937 to predict the placement results.
2. Related Work

We used some famous psychological scales as explanatory variables in machine learning. The following scales are thought to affect academic performance. Deci and Ryan (1985, 2002) studied intrinsic motivation in human behavior. They defined intrinsic motivation as the life force or energy for the activity and for the development of the internal structure. The degree of self-efficacy affects the efficiency of that behavior. According to Bandura (1997), self-efficacy expectancies determine the initial decision to perform a behavior, the effort expended, and persistence in the face of adversity. Sherer et al. (1982) developed a self-efficacy scale.

Task value is a scale focusing on the value aspect of motivation. According to Eccles and Wigfield (1985), task value is divided into three subscales (interest value, attainment value, and utility value). Moreover, Ida (2001) divided task further divided attainment value and utility value into two for a total of five subscales. Attainment value is divided into private attainment value, which positions him/herself with absolute standards by individuals, and public attainment value, which focuses on attention to superiority/inferiority with others. Utility value is divided into institutional utility value, which is used when learning is necessary to pass an examination for employment or admission, and practical utility value, which is used when learning is useful in occupational practice. Ida (2001) also proposed a task value evaluation scale.

According to Duckworth, and Quinn (2009), self-control is needed to achieve goals that require long-term effort. Self-control allows one to focus on a goal (Consistency of Interest) and persevere through difficulties (Perseverance of Effort). They called this combination Grit, and developed an evaluation scale.

Goal orientation is divided into three subscales: mastery orientation, performance approach, and performance avoidance. Elliot and Church (1997) examined their influences and factors. Multi-dimensional competitiveness is divided into three subscales: Instrumental Competitiveness, Avoidance of Competition, and Never-Give-Up Attitude. Ryckman, Hammer, Kaczor, & Gold (1990, 1996), Smither & Houston (1992), and Ota (2010) developed a multi-dimensional competitiveness. Specific questions based on these scales are shown in section 3.2.1.

Some studies investigated these psychological scales and learning. For example, Robbins et al. (2004) examined the relationship between psychosocial and study skill factors (PSFs) and college outcomes. They found that the best predictors for grade point average (GPA) are academic self-efficacy and achievement motivation. Shen, Chen, & Guan (2007) investigated the potential influence of mastery goal, performance-approach, and avoidance-approach goals, individual interest, and situational interest on students' learning in a physical education. They reported that a mastery goal is a significant predictor for the recognition of situational interest.

Machine learning has been used in various fields, including education. In this paper, we use classification machine learning. For example, Sohsah, Guzey, and Tarmanini (2016) classified educational materials in low-resource languages with machine learning. Márquez-Vera, et al. (2016) predicted school dropout rates of high school students at different steps in the course to determine the best indicators of dropout.

3. Method

We used machine learning (supervised learning) to predict the student placement results for a Java programming class in Waseda University. Three explanatory variables were employed: 1. Psychological Test, 2. Programming Task, and 3. Class Questionnaire. Then we evaluated an effective algorithm and the explanatory variables. The results were used to create and evaluate a model. We used a Python library called malss (https://github.com/canard0328/malss/) for the machine learning.

This paper investigated the following research questions (RQs):

- RQ1: How much does each explanatory variable predict the placement results?
- RQ2: What is the best combination of explanatory variables to predict the placement results?
Table 1: Psychological questions.

<table>
<thead>
<tr>
<th>#</th>
<th>Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I like programming.</td>
</tr>
<tr>
<td>2</td>
<td>I am good at programming.</td>
</tr>
<tr>
<td>3</td>
<td>I feel learning to program is interesting.</td>
</tr>
<tr>
<td>4</td>
<td>Programming is necessary for my desired job / advancement examination.</td>
</tr>
<tr>
<td>5</td>
<td>Programming is useful for desired job / advancement examination.</td>
</tr>
<tr>
<td>6</td>
<td>Programming is necessary for practice in my desired occupation.</td>
</tr>
<tr>
<td>7</td>
<td>Programming is useful in my desired occupation.</td>
</tr>
<tr>
<td>8</td>
<td>I think that learning to program helps me grow as a person.</td>
</tr>
<tr>
<td>9</td>
<td>I think that other people respect those who are proficient at programming.</td>
</tr>
<tr>
<td>10</td>
<td>I think that to learn programming can be bragging.</td>
</tr>
<tr>
<td>11</td>
<td>Setbacks don’t discourage me.</td>
</tr>
<tr>
<td>12</td>
<td>I am diligent.</td>
</tr>
<tr>
<td>13</td>
<td>I finish whatever I begin.</td>
</tr>
<tr>
<td>14</td>
<td>I am a hard worker.</td>
</tr>
<tr>
<td>15</td>
<td>I often set a goal but later choose to pursue a different one.</td>
</tr>
<tr>
<td>16</td>
<td>I have difficulty maintaining my focus on projects that take more than a few months to complete.</td>
</tr>
<tr>
<td>17</td>
<td>New ideas and projects sometimes distract me from previous ones.</td>
</tr>
<tr>
<td>18</td>
<td>I am obsessed with a certain idea or project for a short time but later lose interest.</td>
</tr>
<tr>
<td>19</td>
<td>I want to learn to improve my abilities.</td>
</tr>
<tr>
<td>20</td>
<td>I want to learn new things and increase my knowledge.</td>
</tr>
<tr>
<td>21</td>
<td>I want to learn more so others do not think poorly of me.</td>
</tr>
<tr>
<td>22</td>
<td>I want to learn properly so as not to give bad results to those around me.</td>
</tr>
<tr>
<td>23</td>
<td>I learn to improve the results of the tests and evaluations compared to the around me.</td>
</tr>
<tr>
<td>24</td>
<td>When learning something, I like to earn better grades and higher evaluations than other people.</td>
</tr>
<tr>
<td>25</td>
<td>By competing, you can enhance your ability.</td>
</tr>
<tr>
<td>26</td>
<td>Competition motivates me.</td>
</tr>
<tr>
<td>27</td>
<td>If it is boring, I compete with other people to make it interesting.</td>
</tr>
<tr>
<td>28</td>
<td>I do not like to compete.</td>
</tr>
<tr>
<td>29</td>
<td>I do not want to compete if possible.</td>
</tr>
<tr>
<td>30</td>
<td>I do not want to lose.</td>
</tr>
<tr>
<td>31</td>
<td>I feel strongly that I do not want to lose.</td>
</tr>
</tbody>
</table>

3.1 Participants

This study included 65 students. They are second-year undergraduate students at Waseda University in Japan enrolled in a Java programming class. This class is equivalent to the CS1 level. After placement test, 50 students were in the advanced course and 15 were in the intermediate course. In this paper, they solved programming tasks, answered a psychological test, and completed a questionnaire about the class. After the placement, 50 students were in the advanced course and 15 students were in the intermediate course.

3.2 Input Data
Table 2: Psychological scales corresponding to each question.

<table>
<thead>
<tr>
<th>#question</th>
<th>Psychological scale</th>
<th>Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intrinsic motivation</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>Self-efficacy</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>Task Values</td>
<td>Interest value</td>
</tr>
<tr>
<td>4-5</td>
<td>Task Values</td>
<td>Institutional utility value</td>
</tr>
<tr>
<td>6-7</td>
<td>Task Values</td>
<td>Practical utility value</td>
</tr>
<tr>
<td>8</td>
<td>Task Values</td>
<td>Private attainment value</td>
</tr>
<tr>
<td>9-10</td>
<td>Task Values</td>
<td>Public attainment value</td>
</tr>
<tr>
<td>11-14</td>
<td>Grit</td>
<td>Perseverance of Effort</td>
</tr>
<tr>
<td>15-18</td>
<td>Grit</td>
<td>Consistency of Interest</td>
</tr>
<tr>
<td>19-20</td>
<td>Goal Orientation</td>
<td>Mastery orientation</td>
</tr>
<tr>
<td>21-22</td>
<td>Goal Orientation</td>
<td>Performance avoidance</td>
</tr>
<tr>
<td>23-24</td>
<td>Goal Orientation</td>
<td>Performance approach</td>
</tr>
<tr>
<td>25-27</td>
<td>Multi-dimensional Competitiveness</td>
<td>Instrumental Competitiveness</td>
</tr>
<tr>
<td>28-29</td>
<td>Multi-dimensional Competitiveness</td>
<td>Avoidance of Competition</td>
</tr>
<tr>
<td>30-31</td>
<td>Multi-dimensional Competitiveness</td>
<td>Never-Give-Up Attitude</td>
</tr>
</tbody>
</table>

We prepared following three materials: 1. Psychological Scale, 2. Programming Task, and 3. Class Questionnaire. Materials 1 to 3 were used as explanatory variables in machine learning.

3.3 Psychological Scales


Table 2 shows the psychological scales corresponding to each question. Question 1 measured intrinsic motivation. Question 2 measured self-efficacy. We used simple typical questions such as “I like ~.”, and “I am good at ~.” Questions 3 to 10 were based on the task value scale (Eccles and Wigfield, 1985). We used question statements developed by Ida (2001). Questions 11 to 18 were based on the Short Grit Scale (Duckworth, and Quinn, 2009). We used question statements developed by Nishikawa (2015). Questions 19 to 24 were based on Goal Orientation (Tanaka, and Yamauchi, 2000). Questions 25 to 31 were based on Multi-dimensional Competitiveness (Ota 2010).

3.4 Programming Task

We conducted programming tasks in every class. We used Aizu Online Judge (AOJ) to conduct these programming tasks. AOJ is one of the most famous Online Judging System in Japan. AOJ has many programming problems. There are various problems from simple problems such as "Hello World" to difficult problems such as ACM-ICPC (https://icpc.baylor.edu/) past problems. When a user submits his or her program source code via the submission form available on the problem sheet, AOJ checks the correctness of the program on the server side. Table 3 lists the IDs and names of the problems used. Additionally, we set the difficulty level for each problem by considering the correct answer rate, etc. A larger number indicates a more difficult level. Moreover, we measured the source code metrics, which students submitted to AOJ. To collect their source codes, we used Nightmare which is a high-level browser automation library written in JavaScript. To measure the metrics, we used Checkstyle, which is a static analysis tool for Java. We could make an automatic measurement program easily which has from 100 to 200 LOC because each library has simple APIs. The maximum values determined by Checkstyle’s default were used to detect if the maximum value was exceeded for the following metrics: 1. Is Solved, 2. LOC, 3. Boolean Expression Complexity, 4. Class Data Abstraction Coupling, 5. Class Fan Out Complexity, 6. Cyclomatic Complexity, 7. Executable Statement Count, 8. Max Len file, 9. Max Len method, 10. Max Line Len, 11. Max Outer Types, 12. Max Param, 13. NCSS Class, 14. NCSS File, 15. NCSS Method, 16. Npath Complexity, 17. Npath Complexity, 18. Too Many Methods.
Table 3: Problem id, name, and difficulty of programming task of AOJ (All problems are available from http://judge.u-aizu.ac.jp/onlinejudge/description.jsp?lang=en?id=ProblemID).

<table>
<thead>
<tr>
<th>Problem ID</th>
<th>Problem Name</th>
<th>Difficulty Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>Hello World</td>
<td>1</td>
</tr>
<tr>
<td>10001</td>
<td>X Cubic</td>
<td>1</td>
</tr>
<tr>
<td>10002</td>
<td>Rectangle</td>
<td>1</td>
</tr>
<tr>
<td>10009</td>
<td>Circle</td>
<td>2</td>
</tr>
<tr>
<td>10010</td>
<td>Simple Calculator</td>
<td>3</td>
</tr>
<tr>
<td>10003</td>
<td>Small Large or Equal</td>
<td>1</td>
</tr>
<tr>
<td>10004</td>
<td>Sorting Three Numbers</td>
<td>1</td>
</tr>
<tr>
<td>10005</td>
<td>Print Many Hello World</td>
<td>1</td>
</tr>
<tr>
<td>10006</td>
<td>Print Test Cases</td>
<td>1</td>
</tr>
<tr>
<td>10012</td>
<td>Print Rectangle</td>
<td>1</td>
</tr>
<tr>
<td>10013</td>
<td>Print a Frame</td>
<td>2</td>
</tr>
<tr>
<td>10016</td>
<td>Grading</td>
<td>2</td>
</tr>
<tr>
<td>10019</td>
<td>Sum of Numbers</td>
<td>2</td>
</tr>
<tr>
<td>10017</td>
<td>How many ways?</td>
<td>3</td>
</tr>
<tr>
<td>10021</td>
<td>Finding minimum String</td>
<td>3</td>
</tr>
<tr>
<td>10028</td>
<td>Sort I</td>
<td>3</td>
</tr>
<tr>
<td>0121</td>
<td>Seven Puzzle</td>
<td>4</td>
</tr>
<tr>
<td>0030</td>
<td>Sum of Integers</td>
<td>4</td>
</tr>
<tr>
<td>10014</td>
<td>Print a Chessboard</td>
<td>1</td>
</tr>
<tr>
<td>ITP1_5_D</td>
<td>Structured Program I</td>
<td>1</td>
</tr>
<tr>
<td>10023</td>
<td>Shuffle</td>
<td>2</td>
</tr>
<tr>
<td>10020</td>
<td>Counting Characters</td>
<td>2</td>
</tr>
<tr>
<td>1129</td>
<td>HanafudaShuffle</td>
<td>3</td>
</tr>
<tr>
<td>10031</td>
<td>Search II</td>
<td>3</td>
</tr>
<tr>
<td>1160</td>
<td>How Many Islands?</td>
<td>4</td>
</tr>
<tr>
<td>10026</td>
<td>Standard Deviation</td>
<td>1</td>
</tr>
<tr>
<td>10020</td>
<td>Counting Characters</td>
<td>1</td>
</tr>
<tr>
<td>0011</td>
<td>Drawing Lots</td>
<td>1</td>
</tr>
<tr>
<td>1147</td>
<td>ICPC Score Totalizer Software</td>
<td>2</td>
</tr>
<tr>
<td>1129</td>
<td>Hanafuda Shuffle</td>
<td>2</td>
</tr>
<tr>
<td>2102</td>
<td>Rummy</td>
<td>3</td>
</tr>
<tr>
<td>1173</td>
<td>The Balance of the World</td>
<td>3</td>
</tr>
<tr>
<td>1166</td>
<td>Amazing Mazes</td>
<td>3</td>
</tr>
<tr>
<td>1144</td>
<td>Curling 2.0</td>
<td>4</td>
</tr>
<tr>
<td>1133</td>
<td>Water Tank</td>
<td>4</td>
</tr>
<tr>
<td>1302</td>
<td>Twenty Questions</td>
<td>4</td>
</tr>
</tbody>
</table>

3.5 Questionnaire about the Class

We implemented a questionnaire about the class. Participants completed the questionnaire during the class after the placement test. 4 shows the questions. All questions were evaluated on a seven-level scale. These questions were created based on the end-of-term questionnaire that Waseda University for all classes.
3.6 Assignment Test (Programming Quiz)

Table 4: Questionnaire about the class.

<table>
<thead>
<tr>
<th>#</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Are you satisfied with the contents of the class so far?</td>
</tr>
<tr>
<td>2</td>
<td>How much time do you spend learning class contents outside of the class hours in a week?</td>
</tr>
<tr>
<td>3</td>
<td>Do you try to understand the lesson contents?</td>
</tr>
<tr>
<td>4</td>
<td>Do you understand the contents of this class?</td>
</tr>
<tr>
<td>5</td>
<td>Do you think that class materials are easy to understand?</td>
</tr>
<tr>
<td>6</td>
<td>Do you think that the contents of exercises and homework are difficult?</td>
</tr>
<tr>
<td>7</td>
<td>Do you think that the number of tasks and homework is too much?</td>
</tr>
<tr>
<td>8</td>
<td>Do you think that teachers give classes by grasping the understanding degree of the students?</td>
</tr>
<tr>
<td>9</td>
<td>Are you interested in competitive programming like AOJ and contests?</td>
</tr>
<tr>
<td>10</td>
<td>Do you think that this lesson is meaningful?</td>
</tr>
</tbody>
</table>

Table 5: Examination programming quiz.

<table>
<thead>
<tr>
<th>#</th>
<th>Examination sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Create a program that computes the sum of natural numbers from 1 to 100 and outputs it to the display. Do not use mathematical formulas.</td>
</tr>
<tr>
<td>2</td>
<td>Create a program that calculates the sum of squares from 1 to 100 and outputs it to the display. Do not use mathematical formulas.</td>
</tr>
<tr>
<td>3</td>
<td>Create a program to calculate a sequence of numbers (Fibonacci numbers: F(0)=0, F(1)=1, F(n)=F(n-1)+F(n-2)). However, the program should be terminated when F(n) exceeds 10000.</td>
</tr>
<tr>
<td>4</td>
<td>Create a program to calculate a sequence of numbers (Trivonachs number: T(0)=0, T(1)=0, T(2)=1, T(n)=T(n-1)+T(n-2)+T(n-3)). However, the program should be terminated when T(n) exceeds 10000.</td>
</tr>
<tr>
<td>5</td>
<td>Create a program to generate 1000 Java random numbers with natural numbers between 0 and 100. Display their maximum value, minimum value, and average value.</td>
</tr>
<tr>
<td>6</td>
<td>Create a program that displays the number of bills (10,000 yen, 5,000 yen, 1 thousand yen) and coins (500 yen, 100 yen, 50 yen, 10 yen, 5 yen, 1 yen) needed to pay the amount after entering a certain amount on the keyboard. The solution should use the fewest bills or coins possible.</td>
</tr>
<tr>
<td>7</td>
<td>Create a game to hit an integer, which is generated randomly. When the user inputs a value smaller than the correct answer, display &quot;it is smaller than the correct answer&quot;. When the user inputs a value larger than the correct answer, display &quot;it is larger than the correct answer&quot;. The random number is an integer from 0 to 999. If the user does not answer correctly within 10 times, displayed “Game Over”.</td>
</tr>
<tr>
<td>8</td>
<td>Please indicate the execution result of the following three programs. e.g. for(char c='A'; c&lt;='Z'; c++) System.out.print(c); System.out.print(&quot;n&quot;);</td>
</tr>
</tbody>
</table>

Table 5 shows the examination sentences. The examination result was not used as an explanatory variable for machine learning. However, the result was referred by the teacher for class placement. The quiz time was 90 minutes. Additionally, at this test, teachers asked each student which class he/she wants to be in, advanced class or intermediate class (Hope Class).

3.7 Algorithm Selection

This paper used supervised learning algorithms. Five algorithms were tested to create a better model:

- Support Vector Machine with RBF Kernel (SVM)
- Random Forest (RF)
- Support Vector Machine with Linear Kernel (SVML)
- Logistic Regression (LR)
- Decision Tree (DT)
To evaluate the prediction quality of the model, we implemented **Stratified five-fold cross validation**. First, it divides the data set into five pieces so that each label is divided with the same ratio. One piece is used for testing. The remaining four are used for learning. Cross validation calculates the F-measure with precision and recall, verifying each of five divided data sets as test data five times. We'll use the training data as a test set (a closed test).

### 3.8 Feature Selection

To improve the model, especially to avoid a high variance, we investigated the influence of each explanatory variable. Ineffective variables were excluded. In the psychological test, we converted the answers to the 31 questions into scores (1 to 7 points). Then we calculated the sum of the scores by 15 subscales. Next, we measured the metrics for all the tasks solved by the students. Scores ranked by magnitude of the metrics were used as explanatory variables for machine learning because the number of explanatory variables is enormous if each metric for each problem is used. Moreover, we added the total number of answers, the number of answers per difficulty level [**Number of Solved Tasks (AOJ)**, and **Difficulty Level 1 to 4 (AOJ)**].

Finally, we tried to create a model that improved the evaluation score. First, we used the explanatory variable with the best F-measure. Then we added the explanatory variable with the next best F-measure. This procedure was repeated until the model did not improve.

### 4. Results and Discussion

#### 4.1 RQ1: How much does each explanatory variable predicts the placement results?

Table 6 shows the results. The explanatory variables of the measured metrics show high F-measures. **Self-efficacy** and **interest value** also shows high F-measures. As we expected, the F-measure of **Hope Class** was high. This means that these explanatory variables predict the placement results. However, other F-measures in the psychological scales are not very good. Especially, **Never-Give-Up Attitude**, **Perseverance of Effort**, and **Intrinsic motivation** show very low F-measure. **Questions about the class** (Q1-10) show F-measures which are higher than those of psychological scales, lower than those of measured metrics though Q7 (Amount of exercises and homework) shows the lowest F-measure. About the **task value**, the utility values show higher values than the attainment values. From the programming tasks using AOJ, **Number of Solved Tasks (AOJ)** and **Difficulty Level 2 (AOJ)** can predict the placement result to some degree, while **Difficulty Level 4 (AOJ)** shows a low F-measure.

#### 4.2 RQ2: What is the best combination of the explanatory variables to predict the placement results?

We added explanatory variables one by one until the F-measure no longer improved. The best F-measure has a value of 0.937 with DC using the following explanatory variables: (1) **Class Fan Out Complexity**, (2) **Practical utility value**, (3) **Difficulty Level 4 (AOJ)**, (4) **Difficulty Level 3 (AOJ)**, (5) **Interest value**, and (6) **Never-Give-Up Attitude**. Adding more explanatory variables actually decreases the F-measure. A result of a closed test was 0.97. Table 7 shows the F-measures of each algorithm and the best model. By comparing these results, we found that DC was the best algorithm. Figure 1 (right) shows precision, recall, and F-measure of DC. Figure 1 (left) shows a learning curve of DC. Additionally, Figure 2 shows the learning curves of the other algorithms. By comparing these learning curves, there are large gaps between the training scores and cross-validation scores of DC, SVML, RF, and SVM. It means high variance (over-fitting). Thus, if we use more training samples, it can reduce the effect of over-fitting, and lead to improvements in a high variance estimator. On the other hand, even training score of LR is unacceptably bad. It means high bias (under-fitting). Thus, if we use this algorithm and add more features, it can improve a high-bias estimator.

From the result of RQ 1, we expected the results to contain many explanatory variables based on the measured metrics. However, we did not expect **Never-Give-Up Attitude** and **Difficulty Level 4 (AOJ)** to be included because it shows a low F-measure in the previous section.
Questions in the questionnaire about the class. Numbers in parenthesis indicate the maximum values.

<table>
<thead>
<tr>
<th>F-measure</th>
<th>Algorithm</th>
<th>Explanatory Variable Name</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.700</td>
<td>SVM</td>
<td>Q1</td>
<td>Satisfaction with class</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Q2</td>
<td>Learning time</td>
</tr>
<tr>
<td>0.675</td>
<td>SVM</td>
<td>Q3</td>
<td>Effort to understand the contents</td>
</tr>
<tr>
<td>0.762</td>
<td>SVM</td>
<td>Q4</td>
<td>Comprehension of class contents</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Q5</td>
<td>Ease of understanding class materials</td>
</tr>
<tr>
<td>0.670</td>
<td>SVM</td>
<td>Q6</td>
<td>Difficulty of tasks and homework</td>
</tr>
<tr>
<td>0.648</td>
<td>SVM</td>
<td>Q7</td>
<td>Amount of exercises and homework</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Q8</td>
<td>Teacher’s understanding of students’ level</td>
</tr>
<tr>
<td>0.757</td>
<td>SVM</td>
<td>Q9</td>
<td>Interest in competitive programming</td>
</tr>
<tr>
<td>0.787</td>
<td>SVM</td>
<td>Q10</td>
<td>Whether the class is meaningful</td>
</tr>
<tr>
<td>0.667</td>
<td>Dt</td>
<td>Perseverance of Effort</td>
<td>Long-term efforts to achieve the goals</td>
</tr>
<tr>
<td>0.669</td>
<td>RF</td>
<td>Consistency of Interest</td>
<td>Self-control and ability to focus the goal</td>
</tr>
<tr>
<td>0.669</td>
<td>Dt</td>
<td>Mastery Orientation</td>
<td>Enhance ability</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Performance Avoidance</td>
<td>Superior to others</td>
</tr>
<tr>
<td>0.669</td>
<td>RF</td>
<td>Performance Approach</td>
<td>Avoid situations where one’s incompetence is obvious</td>
</tr>
<tr>
<td>0.700</td>
<td>SVM</td>
<td>Instrumental Competitiveness</td>
<td>Achieve another purpose through competition</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Avoidance of Competition</td>
<td>Avoid competition</td>
</tr>
<tr>
<td>0.653</td>
<td>RF</td>
<td>Never-Give-Up Attitude</td>
<td>Do not want to lose</td>
</tr>
<tr>
<td>0.834</td>
<td>SVM</td>
<td>Interest Value</td>
<td>Gain fulfillment and satisfaction</td>
</tr>
<tr>
<td>0.682</td>
<td>LG</td>
<td>Institutional Utility Value</td>
<td>Must pass the exam for employment or admission</td>
</tr>
<tr>
<td>0.681</td>
<td>Dt</td>
<td>Practical Utility Value</td>
<td>Useful for work and study</td>
</tr>
<tr>
<td>0.669</td>
<td>LG</td>
<td>Private Attainment Value</td>
<td>Improve oneself on an absolute scale</td>
</tr>
<tr>
<td>0.669</td>
<td>RF</td>
<td>Public Attainment Value</td>
<td>Improve oneself on a relative scale</td>
</tr>
<tr>
<td>0.871</td>
<td>LG</td>
<td>Self-efficacy</td>
<td>Confidence of one's own ability</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Intrinsic Motivation</td>
<td>Motivation by curiosity and interest</td>
</tr>
<tr>
<td>0.754</td>
<td>RF</td>
<td>AOJ</td>
<td>Total number of questions answered</td>
</tr>
<tr>
<td>0.661</td>
<td>SVM</td>
<td>Difficulty Level 1 (AOJ)</td>
<td># of answers for level 1 problem</td>
</tr>
<tr>
<td>0.820</td>
<td>RF</td>
<td>Difficulty Level 2 (AOJ)</td>
<td># of answers for level 2 problem</td>
</tr>
<tr>
<td>0.700</td>
<td>SVM</td>
<td>Difficulty Level 3 (AOJ)</td>
<td># of answers for level 3 problem</td>
</tr>
<tr>
<td>0.669</td>
<td>SVM</td>
<td>Difficulty Level 4 (AOJ)</td>
<td># of answers for level 4 problem</td>
</tr>
<tr>
<td>0.828</td>
<td>Dt</td>
<td>isSolved</td>
<td>Rank of AOJ</td>
</tr>
<tr>
<td>0.844</td>
<td>SVM</td>
<td>LOC</td>
<td>lines of code</td>
</tr>
<tr>
<td>0.846</td>
<td>SVM</td>
<td>Boolean Expression Complexity</td>
<td># of &amp;&amp;,</td>
</tr>
<tr>
<td>0.827</td>
<td>Dt</td>
<td>Class Data Abstractio Coupling</td>
<td># of instantiations of other classes</td>
</tr>
<tr>
<td>0.897</td>
<td>SVM</td>
<td>Class Fan Out Complexity</td>
<td># of other classes a given class relies on</td>
</tr>
<tr>
<td>0.859</td>
<td>SVM</td>
<td>Cyclomatic Complexity</td>
<td>Min # of possible paths in through source</td>
</tr>
<tr>
<td>0.864</td>
<td>SVM</td>
<td>Executable Statement Count</td>
<td># of executable statements</td>
</tr>
<tr>
<td>0.88</td>
<td>SVM</td>
<td>Max Len file</td>
<td># of files exceeding the max Loc (2000)</td>
</tr>
<tr>
<td>0.841</td>
<td>SVM</td>
<td>Max Len method</td>
<td># of methods exceeding the max Loc (150)</td>
</tr>
<tr>
<td>0.814</td>
<td>SVM</td>
<td>Max Line Len</td>
<td># of lines exceeding the max characters (80)</td>
</tr>
<tr>
<td>0.868</td>
<td>SVM</td>
<td>Max Outer Types</td>
<td># of types declared at the outer (or root) level in a file (1)</td>
</tr>
<tr>
<td>0.849</td>
<td>SVM</td>
<td>Max Param</td>
<td># of parameters exceeding Max (7)</td>
</tr>
<tr>
<td>0.834</td>
<td>SVM</td>
<td>NCSS Class</td>
<td># of classes exceeding the Max non-comment lines in the class (1500)</td>
</tr>
<tr>
<td>0.865</td>
<td>SVM</td>
<td>NCSS File</td>
<td># of files exceeding the Max commenting lines in a file including all top level and nested classes (2000)</td>
</tr>
<tr>
<td>0.850</td>
<td>SVM</td>
<td>NCSS Method</td>
<td># of methods exceeding the Max non-comment lines in the class (50)</td>
</tr>
<tr>
<td>0.896</td>
<td>SVM</td>
<td>Npath Complexity</td>
<td># of possible execution paths through a function (method)</td>
</tr>
<tr>
<td>0.865</td>
<td>SVM</td>
<td>Too Many Methods</td>
<td># of methods exceeding the Max methods at all scope levels (100)</td>
</tr>
<tr>
<td>0.870</td>
<td>SVM</td>
<td>Hope Class</td>
<td>Class which each student wants to be in.</td>
</tr>
</tbody>
</table>
Table 7: The F-measure of each algorithm at the best score (5-fold nested cross validation).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machine (RBF Kernel)</td>
<td>0.82</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.834</td>
</tr>
<tr>
<td>Support Vector Machine (Linear Kernel)</td>
<td>0.787</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.785</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.937</td>
</tr>
</tbody>
</table>

It is interesting that *Practical utility value* is included. It is thought that these variables performed by combining with the former explanatory variables.

5. Threats to Validity

The questionnaires were conducted after the placement test. This could affect the result. Moreover, the best combination may be a local solution. These are threats to the internal validity.

These results are from one class. If this experiment is repeated with another group or organization, the results may differ. Furthermore, the amount of data is small. These are threats to the external validity.
6. Conclusion and Future Work

Machine learning is used to predict the placement results without a traditional placement examination. The explanatory variables are: Psychological Scale, Programming Task, and Student-answered Questionnaire. The target variable is the Placement Result based on an examination by a teacher. We investigated how these three explanatory variables affect the results. Additionally, we created a classification model with a precision, recall, and F-measure of 0.937.

Additional improvements may be possible. For example, there may be a superior algorithm than the ones used in this study. If our method is expanded in the future, it can be applied other situations such as companies' recruitment and placement.

References

Understanding Support System for Causal Relationship in Historical Learning

Tomoko KOJIRI*, Fumito NATEb & Keitaro TOKUTAKEc

a Faculty of Engineering Science, Kansai University, Japan
b Graduate School of Science and Engineering, Kansai University, Japan
c Hiroo Gakuen Junior and Senior High School, Japan

c School of Environment and Society, Tokyo Institute of Technology, Japan

*kojiri@kansai-u.ac.jp

Abstract: In historical learning, to grasp the causal relationship between historical events and to understand factors that bring about important events are significant for fostering the historical learning. However, some students are not able to find events that have causal relationships. The objective of this paper is to develop a support system for understanding the meaning of a causal relationship and making valid causal relation graphs in the historical learning. When events have a causal relationship, a state change in one event causes the other event. To consider these state changes is critically important to connect historical events. This paper proposes understanding scheme for grasping causal relationships between events by arranging state changes. Our system firstly asks students to arrange state changes of historical people according to the historical events, and then to draw the causal relation graph. The described state changes are indicated on the links in the causal relation graph. By observing the state changes on the links, students are not only able to check whether their causal relation graphs are correct, but also understand that state changes are important for grasping the causal relationships.

Keywords: Causal relationship, historical learning, historical thinking, state change

1. Introduction

Historical thinking skill is a reasoning skill to analyze and explain the historical events (Elder et al. 2012 & Ikejiri et al. 2012). This skill is important for inferring about events that will occur in the future (Lee, 2005 & Parkes, et al., 2014). Such skill brings us intrinsic knowledge for our future activities in many fields, such as politics, environmental issues, or architecture (Abbot et al. 1989). The skill should be learned through the historical learning. However, especially in Japan, students tend to memorize facts in historical learning and historical thinking skill is merely acquired.

Understanding the flow of the history and aware of the important factors that cause significant events in history are cornerstones for improving historical thinking. Important factors are estimated by the historical events that have causal relationships. What to regard important factors are different for individuals, so to discuss the differences is also meaningful. In order to discuss the important factors, one of the authors introduce a learning technique of having his middle school students draw graphs representing causal relationships, which we call a causal relation graph. This causal relation graph takes a form of a concept map (Chang et al. 2002), which basically consists of nodes representing historical events and links representing causal relationships. In implementing this technique, one of authors found that some of his students misconstrued the causal relation graph to signify a chronological time line rather than a causal relationship. The objective of this paper, therefore, is to develop a support system for understanding a causal relationship and drawing a valid causal relation graph in the historical learning.

When events have a causal relationship, a state change in one event causes the other event. This research defines the valid causal relation graph as the graph whose links correspond to the state change. To consider state changes is the key to create the valid causal relation graph. Therefore, this paper proposes understanding scheme for valid causal relationships between events by arranging state changes. Our system firstly asks students to arrange state changes of historical events, and then to draw the causal relation graph. The derived state changes are automatically indicated on the created links in the causal relation graph. By observing the state changes on the links, students are not only able to
check whether their causal relation graphs are correct, but also understand that state changes are important for grasping the causal relationships. Horiguchi et al. also focused on teaching causal relationships of historical events using pseudo-haptics of the tablet device (Horiguchi et al. 2016). However, this research only gives awareness information of incorrectness of causal relationships. In addition, it is difficult to define correct causal relationships, since there are various ways of understanding history. Our system encourages students to grasp the way to derive valid causal relationships, such as to consider state changes, by indicating derived state changes to the links in the causal relation graph.

2. Support of Creating Valid Causal Relation Graph of Historical Events

2.1 Valid Causal Relation Graph

Causal relationship graph represents the causal relationships of historical events and takes the form of a directed graph. Nodes correspond to the historical events and links show the causal relationship. Node at the bottom of the arrow is the cause of the node at the top of the arrow.

Assuming a valid causal relationship, state changes triggered by a cause event (Event A) can explain the emergence of the effect event (Event B) as shown in Figure 1. Therefore, if there are causal relationships between historical events, there is a state change that can explain it. As an example, Figure 2 is the causal relation graph that focuses the outbreak of peasant uprisings during the Muromachi period in Japan (1336-1573). The overview of this history is shown in Table 1 where underlined texts show the events occurred during this period. State changes between historical events in the causal relation graph is shown as messages on links. Consider the impact that the “formation of self-governing villages” had on other historical events. By the “formation of self-governing villages”, the solidarity of the peasants got strengthened, which contributed to the “outbreak peasant uprisings,” so clearly there is a valid causal relationship between “formation of self-governing villages” and “outbreak of peasant uprisings.” On the other hand, the “development of a money economy” was not derived by “strengthening solidarity of the peasants,” so there is no causal relationship between the “formation of self-governing villages” and the “development of a money economy.”

![Figure 1. State change and causal relation](image1)

![Figure 2. Example of causal relation graph with state change of historical text shown as Table 1](image2)
Table 1: Example historical text

| During the mid-Muromachi period, the solidarity of the people strengthened due to the formation of self-governing villages. In addition to it, the rise of a money economy caused the peasants to accumulate debt. This led to frequent peasant uprisings and political demands being made to persons of authority in the Bakufu and palladium. Most of these demands were related to the peasants’ desire for a virtuous government. These frequent uprisings weakened the Muromachi Bakufu, and Muromachi Bakufu was unable to effectively handle the situation. |

2.2 Approach for Creating Valid Causal Relation Graph

In order to support creating valid causal relation graph, students need to recognize the state changes by the historical events. Based on this assumption, this paper proposes the thinking process for creating a valid causal relation graph. Figure 3 shows the thinking process. Firstly, students read the historical text and understand its detail. Secondly, they grasp the state changes of historical people along with the historical events. Finally, they find the historical events that have causal relationships and create the causal relation graph by referring to the state change.

![Figure 3. Thinking process for creating a valid causal relation graph](image)

By following this thinking process, students are able to consider the state changes consciously. However, it is difficult to follow this step for students who are not trained to consider the state changes. In addition, even if they grasp the state changes, some of them are not able to reflect the grasped state changes to the causal relation graph. Therefore, this research proposes the system in which students are able to follow the thinking process in creating a causal relation graph. The system also gives awareness regarding to the state changes in creating causal relation graph. By creating the causal relation graph using the system, students are not only able to create a valid causal relation graph, but also to understand the necessity of grasping state changes in considering the causal relation.

The system provides the form for arranging state changes as state transition map in the step of grasping the state changes of historical people. The state transition map is proposed by our research group in which state changes of historical people along with the historical events can be organized (Kojiri et al. 2015). Figure 4 (a) shows the form of state transition map. In this map, circles on top show the historical events and vertical lines represent the timing that each events have occurred. Rows correspond to the historical people’s state changes. The blue squares correspond to the state change of the historical people after the events have occurred. Using this map, state changes that are occurred after the historical events have been organized.

Our system also provides the environment to draw a causal relation graph by considering the state changes. The state change which is occurred after the cause event is the trigger of the result event as shown in Figure 4 (b). So, our system displays the state changes after the cause event in the state transition map to the links in the causal relation graph so as to make students check the validity of their causal relationships.

3. System for Supporting Creation of Causal Relation Graph

We have developed a system for supporting students to create valid causal relation graph. The system consists of two subsystems: a state transition map generation support system and a causal relation graph generation support system. Figure 5 shows a screenshot of the interface for the state transition map generation support system. By selecting a previously covered history theme in the History Selection Area, a brief historical overview is appeared in the Text Display Area, and the student is allowed to create a state transition map in the State Transition Map Display Area. When the student clicks the intersection of a historical person and an event, the window for inputting the state change is appeared. In
the window, the student needs to input the type of state and selects the type of its change from “UP” and “DOWN.” When the state change is inputted, the blue square describing the state change is appeared in

![Figure 4. Relation between state transition map and causal relation graph](image)

Figure 4. Relation between state transition map and causal relation graph

![Figure 5. The interface for the state transition map generation support system](image)

Figure 5. The interface for the state transition map generation support system

the State Transition Map Display Area. By clicking on the Render Graph Button in Figure 5, the system switches over to the causal relation graph generation support system.

Figure 6 shows a screenshot of the interface for the causal relation graph generation support system. Historical events are displayed in the Historical Event Display Area. The student can draw a causal relationship in the Causal Relation Graph Display Area by left clicking a cause event, right clicking an effect event, and then pushing the Render Button. State changes generated by a cause event are revealed by moving the cursor over the link as shown as Figure 6. Revealing the state changes in this way makes it obvious to the student whether a causal relationship is valid or not.

![Figure 6. The interface for the causal relation graph generation support system](image)

Figure 6. The interface for the causal relation graph generation support system
4. Experimental Trials

We conducted two trials to evaluate the validity of understanding state changes and the effectiveness of the support systems. As for the first trial, 5 university students (A-E) were recruited as subjects. As for the second trial, 8 middle school students (a-h) were subjects. As instructional materials, we have prepared an instructional video and several written passages on the theme “Transition from the Great Depression to World War II.”

First, the students were instructed to read the history instructional materials and to draw a causal relation graph on the paper using a pen (causal relation graph 1). Next, we had them generate state transition maps using the state transition map generation support system. In the system, historical people and historical events of the state transition maps are given in advance. After that, they were allowed to modify their original graphs—causal relation graph 1—to create causal relation graph 2. And finally, the students created causal relation graphs using the causal relation graph generation support system (causal relation graph 3).

Tables 2 shows (i) whether the causal relation graphs were changed before and after generating state transition maps and before and after using the system and (ii) responses of the students on the questionnaire as to whether they had any experience in considering causal relationships in learning history. For (i), Y means that the students changed their causal relation graphs to the valid one and N means that they did not change them. In this experiment, no students changed the graphs to invalid one. For (ii), the students were asked to select one answer from given five options ranging from a definite “1. No” to a definite “5. Yes.”

Table 2: Results: (i) causal relation graph changes and (ii) response of questionnaire

<table>
<thead>
<tr>
<th>University students</th>
<th>Middle school students</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Change of causal relation graph from causal relation graph 1 to 2</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Change of causal relation graph from causal relation graph 2 to 3</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Have you ever considered causal relationships in the study of history?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

According to the result, 7 out of 13 students made changes in total either after generating the state transition maps or after creating the causal relation graph. Among the university students, the change of the causal relation graph was observed after generating the state transition maps, while that was observed after creating the causal relation graph for middle school students. We would infer that such differences depend on whether the students could consider the meaning of deriving the state changes in generating the state transition map or not. That is, the university students inferred why they need to consider the state changes while creating the state transition map, but the middle school students did not. Therefore, the middle school students found the relations between state changes and causal relation when they were creating the causal relation graph.

Now look at 6 students who did not change the causal relation graphs. Among them, students H, c, and e answered 4 or 5 for (ii), which means they had experiences in considering the causal relationships in history. Such students were able to consider the state changes before the experiment and created the valid causal relation graph from the start of the experiment. Therefore, we would infer, that for the students who had never given much thought to causal relationships in the past, the proposed thinking process would be quite effective in helping understand the concept of causal relationships.

Next, we look at the number of created state changes on the state transition map in Table 3. The students A, E, b, and f derived less than 10 state changes. These students did not produce enough state transition maps. Among them, students except E did not change the causal relation graph. This finding would indicate that this thinking process and the system were not effective for students who are not able to derive the enough state changes. This is an issue that we must address in future work that to develop the support method to make students derive the state changes from the history text.
Table 3: The number of derived state changes in the state transition map

<table>
<thead>
<tr>
<th></th>
<th>University students</th>
<th>Middle school students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

5. Conclusions

This study focuses on understanding the meaning of the causal relation in the history and creating valid causal relation graph. It proposes a novel thinking process that organizes state change in a way that helps students understand causal relationships. We also implemented a learning environment in which students are able to follow the proposed thinking process and are aware of the relations between state changes and causal relationships. The experimental trials demonstrated that the activities involved in producing state transition maps were effective in helping university students grasp causal relationships, while the system functions for creating causal relation graphs helped the middle school students understand causal relationships. The difference in responsiveness to the different learning levels of the two groups of student subjects. We also found that students who did not modify the causal relation graphs at all did not produce enough state changes in the state transition maps. For our next step, we need to come up with additional instructional aids for this kind of students who clearly did not understand the concept of state change.

In addition, we need further experiments to evaluate the effectiveness of the thinking process and the support system, since the number of the subjects in the experimental trials was small.

Acknowledgements

The work was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) (No. 16H03089, 15H02934).

References

Extraction of Relationships between Learners’ Physiological Information and Learners’ Mental States by Machine Learning

Yoshimasa TAWATSUJI*, Tatsuro UNO, Keita OKAZAKI, Siyuan FANG & Tatsunori MATSUI

*Faculty of Human Sciences, Waseda University, Japan
bSchool of Human Sciences, Waseda University, Japan
cComprehensive Research Organization, Waseda University, Japan
dGraduate School of Human Sciences, Waseda University, Japan

Abstract: The estimation of learners' mental states during the interaction between teachers and learners is a very important problem in improving the quality of teaching and learning. In this experimental study, we developed a deep learning neural network (DLNN) system that extracted the relationships between a learner’s mental states and a teacher's utterances plus the learner's physiological information. The learner's physiological information consisted of the NIRS signals, the EEG signals, respiration intensity, skin conductance, and pulse volume. The learner's mental states were elicited through the learner's introspective reports using the Achievement Emotions Questionnaire (AEQ). According to the AEQ, the learner’s mental states were divided into nine categories: Enjoy, Hope, Pride, Anger, Anxiety, Shame, Hopelessness, Boredom, and Others. In a simulation, the DLNN system exhibited the ability to estimate the learner’s mental states from the learner’s physiological information with high accuracy.

Keywords: Intelligent Mentoring System, physiological data, deep learning, mental state estimation

1. Introduction

1.1 Research Background and Objective

When evaluating the effects of teaching and learning, knowing learners’ mental states during their learning processes is vital. Many fundamental studies in educational technology have investigated the relationships between learners’ physiological information, such as eye movements and sweating, and learners’ behaviors and mental states. Recently, with computers and biometric instruments becoming increasingly higher in performance and lower in cost, collecting a great amount of physiological data and processing them efficiently in real time has become easy. Therefore, many education support systems have been developed that can automatically estimate learners’ mental states from learners’ behavioral information and physiological information, to a certain extent. On the other hand, it is generally acknowledged in the field of pedagogy that the interactions between teachers and learners can influence learners’ mental states. Therefore, it is interesting to formalize the relationships between teachers’ behaviors and/or utterances and learners’ mental states (or the factors related to learners’ mental states). It is expected that the performance of the education support systems can be improved by integrating the formalized knowledge into the modules of the systems that serve to estimate learners’ mental states. In our previous study, we tried to formalize the relationships between learners’ mental states and teacher’s utterances plus learners’ physiological information collected during teacher–learner interactions (Takehana and Matsui, 2016). We used the technique of association rule mining to process the teachers’ utterances, the learners’ physiological information, and the learners’ introspective reports. On the other hand, Horiguchi, Kojima, and Matsui (2010), Kojima, Muramatsu, and Matsui (2014), and Fujiyoshi, Yoshimura, Kunze, and Kise (2015) suggested the possibility of applying machine leaning
techniques to the development of education support systems. Therefore, in the present study, we employed the technique of deep learning to analyze the data on teacher-learner interactions for the purpose of developing a system that can automatically estimate learners’ mental states.

1.2 Synopsis

Section 2 describes the experiment that we conducted to obtain the empirical data that we used to train and test our machine learning systems. In this experiment, we recorded the teacher’s utterances and the learners’ physiological information and mental states. Section 3 introduces the construction of the first machine learning system by using a three-layer artificial neural network (3LNN). The inputs of the 3LNN are the data of a teacher’s utterances and a learner’s physiological information. The output of the 3LNN is the learner’s mental states. The 3LNN was trained and tested using the experimental data. In Section 4, we redefined the inputs and output of the NN system as time-series data, and increased the number of hidden layers, producing a deep learning neural network (DLNN). The experimental data were used to train and test the DLNN. Section 5 describes the results of the construction of the two NN systems and discusses their significance. Section 6 summarizes this study and introduces some implications for future work.

2. Collection of Multifaceted Learning-Related Data

To collect multifaceted learning-related data, we conducted a biometric experiment. The subjects were one teacher and one learner of a private tutoring school. The learner was a junior high school student taking extra classes in the private tutoring school. We obtained the informed consents of participation from the student and the guardians of the student through the teacher. The physiological information collected in the experiment included near-infrared spectroscopy (NIRS) signals (recorded using Hitachi WOT-100), respiration intensity, skin conductance, and pulse volume (recorded using NeXus). The subject put on all the above instruments when taking a regular class as the experimental task. To align the recording time of these instruments, we placed time markers at the beginning and end of the measurement. The NIRS data were treated using the global average reference method (Nozawa and Kondo, 2009; Hirayama, Watanuki and Kaede, 2012). The entire experiment was recorded by three video cameras set at different locations in the classroom. Based on the video content, we divided the teacher’s utterances into nine categories. We then revised part of the categories by referring to the categorization proposed by Fujie (2000), Shimizu and Uchida (2001), and Kishi and Nojima (2006). The final version of our categories is “Explaining,” “Questioning,” “Comprehension Checking,” “Repeating,” “Praising,” “Task Fulfillment Checking,” “Alerting,” “Chatting,” and “Others.” Using these nine categories, we labeled the teacher’s utterances while watching the video. After several days, the subject was asked to report the changes in the mental states during the course of the experiment while the subject watched the video. The Achievement Emotions Questionnaire (AEQ) (Pekrun, Goetz, Frenzel, Barchfeld and Perry, 2011) was used to divide the mental states into nine categories: “Enjoyment,” “Hope,” “Pride,” “Anger,” “Anxiety,” “Shame,” “Hopelessness,” “Boredom,” and “Others.” The annotation of the video using the nine mental state categories was carried out on a computer program that we developed. The program allowed the subject to select from among nine buttons that represent the nine mental state categories.

3. Estimation of a Learner’s Mental States Using a Three-Layer Neural Network

This section introduces the 3LNN that we constructed in the first try in this study to automatically estimate the learner’s mental states from the learner’s physiological information that we collected during the experiment. We selected a 63-s segment (19 min 37 s–20 min 40 s) from the 60-min video as the object of the data analysis because this segment contained frequent teacher–learner interactions.

3.1 Data Structure and Neural Network Architecture
The inputs of the 3LNN are 1) the learner’s cerebral blood flow measured by NIRS signals (5Hz), 2) the learner’s respiration intensity (32Hz), 3) the learner’s skin conductance (32Hz), and 4) the teacher’s utterances (labeled by the five categories “Explaining,” “Questioning,” “Comprehension Checking,” “Alerting,” and “Others”). To regulate the granularity of the input data types 1–3, linear interpolation was applied to the coarse-grained data types according to the granularity of the data type with the finest granularity. The result of this granularity regulation is a 2024*5 data matrix. The five columns represent respectively the five input data types and the learner’s mental states. Then, the input data of types 1–3 were normalized to zero mean and unit variance. The 3LNN consists of one input layer, one hidden layer, and one output layer. The number of units in the hidden layer was determined though simulations. In the simulations, we tested the number of hidden-layer units from one to 25, and found that the number 19 had the minimum learning loss and the maximum accuracy. The results of the simulations are shown in Figure 1.

3.2 Methods and Results of System Performance Simulation

A simulation for testing the performance of the 3LNN was run on Python 3.5 with TensorFlow (ver 0.12.1). The activation function of the hidden layer was the tanh function, and the activation function of the output layer was the softmax function. The cost function was the cross-entropy error function. The gradient descent was used as the optimization method. The learning rate was set to 0.05. For cross-validation, 60 percent of the experimental data (totally 2049 data sets) were used as the training data, and the remaining 40 percent were used as the validation data. The learning phase had 5000 iterations. Part of the results of the learning phase is displayed in Figure 2. Figure 2 shows that the learning process converged at an early stage. We performed the cross-validation ten times and the accuracies obtained were 0.900, 0.917, 0.910, 0.917, 0.910, 0.914, 0.915, 0.899, 0.912, and 0.905. This implies that the 3LNN can estimate the learner’s mental states from the learner’s physiological information with high accuracy.

![Figure 1. Changes in learning loss and accuracy that resulted from the adjustment of the number of the hidden-layer units.](image1)

![Figure 2. Results of the system performance simulation of the 3LNN.](image2)
4. Mental State Estimation Using Deep Learning

In the 3LNN described in Section 3, the temporal aspects of the input and output data were not taken into account. Only the relationships between the physiological information and the mental states at a single time point were discussed. Therefore, our next step is to increase the number of hidden layers in a try to improve the estimation accuracy of the NN system.

4.1 Real-Time Data Processing

As described in Section 3, we used a 3LNN to extract the relationships between the learner’s physiological information and mental states that transpired at a single time point. Problems with regard to the following two points remain. First, from the perspective of application, normalizing the input data is difficult. More specifically, when real-time data processing is required, e.g., when developing an Intelligent Mentoring System (IMS), it is hard to apply the global average reference method to the NIRS data and normalize the physiological data. Hence, in the case of IMS development, using the original NIRS data and the original physiological data is necessary. Second, the learner’s mental state at a certain time point \( t \) may influence the learner’s physiological information and the teacher’s utterances. This indicates the possibility that the learner’s physiological data and the teacher’s utterances after \( t \) may also contain the information about the learner’s mental states at \( t \). This time lag was not considered in the construction of the 3LNN. Hence, we wish to integrate this time lag in the NN construction.

![Figure 3. Architecture of the DLNN.](image)

4.2 Data Structure and Neural Network Architecture

The input and output data of this deep learning neural network (DLNN) were the same as those of the 3LNN. However, the physiological data were not normalized in the case of the DLNN. The DLNN consists of one input layer, two hidden layers, and one output layer. The input data are composed of the input data types 1-4 that range from the time point \( \tau - t \) to the time point \( t \), and the output data are the mental state at the time point \( \tau - t \). In this way, the time lag was integrated in the input and output data structures of the DLNN (shown in Figure 3). As in the case of the 3LNN, the number of the units in the two hidden layers was determined through simulations. In the simulations, we gradually changed the numbers of the hidden-layer units, and found that the DLNN had the best performance in terms of learning loss and accuracy when the first hidden layer had 19 units and the second hidden layer had 20 units. Hence, we set the number of the units in the first hidden layer to 19, and that of the second hidden layer to 20. Figure 4 shows how learning loss and accuracy change when the number of the units in the first hidden layer changes from 10 to 19, with the number of the units in the second hidden layer fixed at 20. Besides, time lags of nine different lengths were tested in the simulations; that is, \( \tau \) took the values 2, 3, 4, 5, 6, 7, 8, and 9. The results of the simulation show that irrespective of the manner in which we adjusted the number of the units in the first hidden layer, the accuracy always reached its maximum when \( \tau \) was 9. Hence, in the system performance simulation (described in Section 4.3), \( \tau \) was set to 9.
4.3 Methods and Results of System Performance Simulation

The performance of the DLNN was tested through a simulation run on Python 3.5 with TensorFlow (ver 0.12.1). The activation function of the two hidden layers was the tanh function, and that of the output layer was the softmax function. The cost function was the cross-entropy error function. The gradient descent was used as the optimization method. The learning rate was set to 0.05. As explained in Section 4.2, the number of the units in the first hidden layer was set to 19, and that of the second hidden layer was set to 20. The cross-validation was conducted ten times. For each time, 60 percent of the experimental data were used as the training data, and the remaining 40 percent were used as the validation data. The learning phase, which sought the minimum cross-entropy, contained 5000 iterations. Part of the results of the learning phase is shown in Figure 5. Figure 5 shows that the learning process converged at an early stage. The accuracies achieved during the cross-validations were 0.799, 0.842, 0.797, 0.793, 0.764, 0.811, 0.789, 0.829, 0.768, and 0.808 (mean = 0.8; sd = 0.023). This implies that the DLNN can estimate the learner’s mental states from the learner’s physiological information with an accuracy of approximately 80% even without data standardization.
5. Discussion

We experimentally verified that the DLNN had the ability to estimate the learner’s mental states from the learner’s physiological information with an accuracy of approximately 80% even without data standardization. One reason for this high performance may be that in the simulation, the possible estimation results were limited to only five mental state categories. Another point to note is that in the simulation, the learning processes converged at a very high speed. This suggests the risk of over-fitting.

6. Summary and Implications for Future Work

In this study, we developed a machine learning system that could automatically estimate a learner’s mental states from the learner’s multifaceted learning-related information, to a large extent. The physiological information was recorded in an experiment using biometric instruments. We constructed a DLNN to extract the relationships between the learner’s physiological information and the learner’s mental states, and carried out a simulation to test the performance of the DLNN. The results of the simulation suggest that the DLNN has the ability to estimate the learner’s mental states from the learner’s physiological information with an accuracy of approximately 80% even without data standardization. Nevertheless, because in the simulation the training and test data contained only a small number of mental state categories, there exists the possibility of over-fitting. To solve this problem, in our future studies, we plan to use the data of the entire video, instead of the 63-s segment that was used in this study. Another way to improve the system performance is to find a better method to optimize the number of the hidden layers and the numbers of the units in each hidden layer. In addition, it is important to understand the relationships extracted by the DLNN in the domains of educational technology and pedagogy. Therefore, in future we will try to read what the hidden layers have learned by analyzing and visualizing the weights in the DLNN.

References

Predicting Student Carefulness within an Educational Game for Physics using Support Vector Machines

Michelle P. BANAWAN\textsuperscript{a,b}, Ma. Mercedes T. RODRIGO\textsuperscript{b} & Juan Miguel L. ANDRESc

\textsuperscript{a}Ateneo de Davao University, Davao City, Philippines
\textsuperscript{b}Ateneo de Manila University, Quezon City, Philippines
\textsuperscript{c}University of Pennsylvania, Philadelphia, USA

*mpbanawan@addu.edu.ph

Abstract. Student carefulness is defined as being attentive, mindful or focused on the task at hand. In this paper, we create a predictive model for student carefulness within an educational game called Physics Playground (PP). We used game logs and manually-labeled gameplay clips of 54 students from the Philippines to develop three support vector regression models that predict carefulness using: (1) predictors of the game developers, (2) predictors from social science research, and (3) the combination of these predictors. After preprocessing and feature selection, the support vector regression models were able to significantly predict student carefulness. This research’ empirical findings suggest that carefulness in Physics Playground can best be predicted by expanding the model of the game developers and including predictors that have been previously researched in the broader social science literature.

Keywords: carefulness · machine learning · support vector machines · regression · Physics Playground

1. Introduction

Carefulness is a characteristic of action that involves giving cautious attention, being thorough, alert, attentive, heedful or mindful. Social science and education researchers have extensively studied student carefulness. They have linked carefulness to improved problem-solving skills and higher-order thinking (Whimbey, 1980). On the other hand, students who are not careful have been found to be prone to impulsive, hurried, or incomplete problem solving (Kirby, Winston, and Santiesteban, 2005) and to making mistakes even after a skill has already been mastered (San Pedro, Baker, and Rodrigo, 2011). When a student is being careful, he/she is most likely to have self-discipline and avoid trivial errors (Gong, Beck, and Heffernan, 2009).

We study carefulness within one such learning environment, a game for Newtonian Physics called Physics Playground (PP). The creators of Physics Playground developed a model for student carefulness within PP. These predictors include: number of objects drawn, number of object limits reached in the game, average time in an attempt, average time before making an action in the first attempt, and average time between actions. However, they have not empirically validated this model. Further, the model does not include indicators of novelty, mastery, reflectivity, and other constructs that social science literature has cited as related to carefulness.

2. Research Goal

The goal of this study is to create a predictive model for carefulness by empirically validating the carefulness predictors of PP developers and expanding that model to include additional features from social science constructs that the Physics Playground developers may have not considered.
3. Physics Playground

Physics Playground (Shute et al., 2013) is a two-dimensional educational game that was designed to help secondary school students understand the application of Newton’s three laws of motion. The students are required to bring the green ball to a red balloon by drawing agents like the ramp, pendulum, lever and springboard. Shute et al. modeled carefulness as an “unobservable” or a construct and was mapped to a number of “observables” or indicators.

4. Social Science Research Findings on Carefulness

The broader social science literature has found a number of cognitive and non-cognitive constructs that are related to carefulness. Examples of these will be briefly discussed in this section.

Novelty is an example of these constructs. It is defined in the Merriam-Webster dictionary as the quality or state of being new. Prior work (Ostafin and Kassman, 2012) has found that novelty is related to a student’s problem solving behavior such that when a student is given new problems where he/she can be more creative and give non-routine solutions or answers, he/she seemed to exercise more care in solving the problem. Within PP, student actions that might be indicative of novelty include attempts on new problems and use of new solutions to previous problems attempted. Another construct that has been frequently related to carefulness is mastery. Mastery is the possession of a skill or knowledge. It occurs when the content and objectives of instruction are learned (Bloom, 1968). There has been a number of studies that investigated the relationship that exist between carefulness and mastery, e.g. the student’s level of proficiency in solving problems and degree of carefulness (Tsiriga, Virvou, 2002), student’s intellectual performance and mindful engagement in games and computerized learning platforms (Salomon, Perkins, and Globerson, 1991), etc. Within PP, a student action or event that might be indicative of mastery would be his correct solutions to problems, i.e gold and silver badges, as PP awards these badges to students every time that they are able to solve a problem. Reflectivity or having the quality of being reflective is described as thinking carefully. It is another construct that has been investigated by researchers interested in student carefulness. It is likened to mindfulness or attentiveness and was seen to help students in problem solving and in academic performance. Its characteristics include being aware and attentive to the present and immediate experience and involves actions like re-reading the problem, backtracking, understanding the problem better, reviewing, rechecking and ensuring that everything has been considered. Related work show findings that reveal relationships between students’ habitual action and lower levels of reflection (Lim, 2011), and an investigation of the nature of reflectivity and its relationship to learning goals (Lin, et al., 1999). Within PP, a student action that might be indicative of reflectivity would be when the student is taking time in solving the problems correctly.

5. Support Vector Machines for Prediction

Support vector regression has been used for small data sets of educational data to find “support vectors” that allow the widest margin between classes. It tolerates outliers and collinearity in data better than linear regression (Ashlay, Chan and Ikeda, 2006) and addresses the pre-requisite assumptions of normality and linearity for regression models. It is known to be able to get good generalization even with a limited number of learning patterns (Basak, Pal, and Patranabis, 2007). Support vector machines transforms the original data points in the training data to a higher-dimensional feature space and a linear regression computation is performed in this high dimensional feature space (Guajardo, 2006). Support vector machines aim to minimize the classification error of both training data and unseen data and has been known to outperform conventional classifiers, most especially when the number of training data is small (Abe, 2005).
6. Data Collection and Processing

Primary data was gathered enlisting the participation of 180 high school students from 3 universities in the Philippines. These students were given 120 minutes to play the game, during which the game logged all player interactions. PP automatically recorded player events/actions. These events have been captured in the logs resulting to an entry in the logs for (roughly) every second of gameplay. These raw logs were parsed to produce a comma separated value (csv) file which we later used in further summarization and preprocessing. A single attempt (on a single problem) is a summary of information of all the rows between the Level Start action to the Level End action from the interaction logs. On average, an attempt consisted of 4,373 actions (rows) with the shortest attempt only having 11 actions (rows) and the longest attempt having 109,091 actions (rows).

Parallel to the automated capture of the interaction logs, student usage/gameplay is also recorded such that they can be viewed as videos using PP’s Replay Viewer utility. We divided these videos into clips that correspond to attempts. Unfortunately, not all attempts had equivalent gameplay videos. There were only 2,640 playable clips, all in all. A stratified sampling of these clips (1,990) was selected, video-casted to mp4 format and a utility was developed for the ground truth labeling. Physics experts/professors have been consulted to come-up with the criteria for labeling the clips. The coders, then, referred to these criteria during the coding process. The clips were then coded via consensus as to 1-least careful, 2-somewhat careful, and 3- very careful. Clips that cannot be coded as such were given a code of 0 for undetermined. Ground-truth coding used the experts’ criteria and were not based on any of the features of Shute et al. nor candidate features of mastery, novelty or reflectivity. Finally, the attempts were further summarized and aggregated to student level details to build models based on Shute, et al.’s (2013) predictors.

7. Findings

7.1 Feature Engineering

We characterized each student with the following features:

1. number of object limits reached in problem [ObjectLimits]
2. number of objects drawn per level [ObjectsDrawn]
3. average time in seconds spent drawing per level [TimeDrawing]
4. average time in seconds between actions [TimeActions]
5. average time in seconds before making an action on the first attempt [Time1stAttempt]
6. number of gold badges earned [CountOfGold]
7. number of silver badges earned [CountOfSilver]
8. average difference between the time spent on problems and the median time for the problem/level [timediff]
9. count of unique solutions [UniqueSolutions]

Features 1-5 were taken from Shute, et al. (2013). Features 6-9 were candidate predictors indicative of other constructs that have been found to be related to carefulness. For this paper, we are trying to capture the unique attempts by looking at the problems that the students answered for the first time. However, even if such attempt is not the first attempt yet the student used a solution that is different from the previous solution used (e.g. using a springboard when the previous solution was a pendulum), then the attempt is still considered as novel. Initially two derived features were extracted for this purpose: (1) unique problems attempted and (2) unique solutions to non-unique problems, both having equal weights. These two features were later combined into a single feature, i.e. unique solutions as both refer to the same characteristic. The formula used was unique solutions = (unique problems attempted + unique solutions to non-unique problems) / number of attempts, such that a perfect score of 1 means that the student had all unique solutions or attempted all problems only once.

Further, in PP, a student is able to earn badges every time he/she solves a problem. A gold badge is earned if the student has drawn three, or less, objects in coming up with the solution
and a silver badge is earned if the solution entailed drawing more (greater than three) objects. It is when a student draws more objects than necessary that he/she reaches the set object limits for the problem (feature #1). Hence, we also included the count of gold badges and count of silver badges as candidate predictors. We also felt that the time spent on solving the problems should also be relative to the overall time-based performance on a specific problem/level, i.e. median time for the problem/level. We, then, engineered another time-based feature - timediff, to be able to see if an attempt took longer than the median time or took less than the median time for the specific problem/level. A high timediff value means that the student is slower than other students, and a low timediff value means that the student is faster.

7.2 Descriptive Summary of the Features

Given 9 features of 54 labeled student instances, descriptive statistics are computed (table 1).

Table 1: Descriptive Statistics of the Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Limits</td>
<td>0</td>
<td>55</td>
<td>5.61</td>
<td>9.51</td>
</tr>
<tr>
<td>Objects Drawn</td>
<td>349</td>
<td>2,201</td>
<td>832.13</td>
<td>350.18</td>
</tr>
<tr>
<td>Time Drawing</td>
<td>1,930.10</td>
<td>7,975.70</td>
<td>4,117.66</td>
<td>1,126.82</td>
</tr>
<tr>
<td>Time Actions</td>
<td>88,280.40</td>
<td>326,090.63</td>
<td>169,168.73</td>
<td>49,815.30</td>
</tr>
<tr>
<td>Time 1st Attempt</td>
<td>9,853.55</td>
<td>41,804.53</td>
<td>19,781.12</td>
<td>8,235.60</td>
</tr>
<tr>
<td>timediff</td>
<td>-73,883.50</td>
<td>219,703.03</td>
<td>18,457.63</td>
<td>53,319.30</td>
</tr>
<tr>
<td>Count Of Silver</td>
<td>7</td>
<td>34</td>
<td>17.80</td>
<td>6.93</td>
</tr>
<tr>
<td>Count Of Gold</td>
<td>0</td>
<td>14</td>
<td>5.62</td>
<td>3.38</td>
</tr>
<tr>
<td>Unique Solutions</td>
<td>15</td>
<td>76</td>
<td>31.59</td>
<td>11.86</td>
</tr>
<tr>
<td>Carefulness</td>
<td>1.26</td>
<td>2.71</td>
<td>2.09</td>
<td>0.28</td>
</tr>
</tbody>
</table>

The average carefulness label of the students ranged from 1.26 (1.0 – least careful) to 2.71 (3 – very careful), with an average of 2.09 (standard deviation of 0.28) which shows that, at average, students were somewhat careful during gameplay.

7.3 Feature Selection

Before building the models, we forward-selected the most efficient set of attributes per dataset, i.e. dataset with Shute et al.’s features only, dataset with candidate features from social science literature and the dataset combing all the features The remaining feature subsets are shown in table 2.

Table 2: Feature subsets of the three datasets after Forward Selection

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset with Shute, et al.’s features</td>
<td>Objects Drawn</td>
</tr>
<tr>
<td></td>
<td>Time 1st Attempt</td>
</tr>
<tr>
<td>Dataset with candidate features from Social</td>
<td>Unique Solutions</td>
</tr>
<tr>
<td>Science literature</td>
<td>Count Of Gold</td>
</tr>
<tr>
<td></td>
<td>Count Of Silver</td>
</tr>
<tr>
<td>Dataset with combined features</td>
<td>Unique Solutions</td>
</tr>
<tr>
<td></td>
<td>Count Of Gold</td>
</tr>
<tr>
<td></td>
<td>Count Of Silver</td>
</tr>
<tr>
<td></td>
<td>Time 1st Attempt</td>
</tr>
</tbody>
</table>

Table 2 shows that only the number of objects drawn (Objects Drawn) and average time before making an action on the 1st attempt (Time 1st Attempt) were selected as features out of the 5 proposed predictors of Shute, et al. For the candidate features (from Social Science literature), three out of four features were selected, only the difference between the time spent on the attempt and the median time for that level (timediff) was not selected. For the combined dataset, the
features selected were: the count of unique solutions (UniqueSolutions), number of gold and silver badges (CountOfGold and CountOfSilver), and the average time spent before making an action on the first attempt.

### 7.4 Carefulness SV Regression Models

With the outliers removed and the features selected through forward selection, the support vector machine algorithm was used resulting to the three SV Regression models in table 3:

<table>
<thead>
<tr>
<th>Model</th>
<th>Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>ObjectsDrawn</td>
<td>-0.062</td>
</tr>
<tr>
<td></td>
<td>Time1stAttempt</td>
<td>-0.008</td>
</tr>
<tr>
<td>Model 2</td>
<td>UniqueSolutions</td>
<td>-0.196</td>
</tr>
<tr>
<td></td>
<td>CountOfGold</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td>CountOfSilver</td>
<td>0.117</td>
</tr>
<tr>
<td>Model 3</td>
<td>UniqueSolutions</td>
<td>-0.195</td>
</tr>
<tr>
<td></td>
<td>CountOfGold</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>CountOfSilver</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>Time1stAttempt</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Table 3: SV Regression Models Weight Vector

From Model 1, only the number of objects drawn (ObjectsDrawn) and average time before making an action on the 1st attempt (Time1stAttempt) came out as the significant predictors. Both predictors are negatively weighted. For this model, student carefulness can be attributed to and predicted by both ObjectsDrawn and Time1stAttempt.

For Model 2, CountOfGold and CountOfSilver are positively weighted significant predictors, and UniqueSolutions is a negatively weighted significant predictor.

For Model 3, CountOfGold, CountOfSilver and Time1stAttempt are positively weighted significant predictors and UniqueSolutions remained to be a negatively weighted significant predictor, consistent with Model 2. Further, the novelty of the problems attempted were not predictive of carefulness, as we initially suspected, and were predictive of non-carefulness instead. It is interesting to note that contrary to Model 1, the sign of Time1stAttempt in this expanded model has changed from negative to positive, which implies that the time that the student takes before making an action on the first attempt when combined with the badges earned predicts carefulness. Unlike in Model 1 where the time that the student takes before making an action on the first attempt predicts non-carefulness when combined with the number of objects drawn.

To evaluate the models, 10-fold cross validation was used resulting to the following average performance vector (table 4):

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>Correlation (R)</th>
<th>SquaredCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.244 +/- 0.142</td>
<td>0.345 +/- 0.362</td>
<td>0.250 +/- 0.332</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.162 +/- 0.057</td>
<td>0.688 +/- 0.247</td>
<td>0.534 +/- 0.224</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.159 +/- 0.065</td>
<td>0.652 +/- 0.309</td>
<td>0.521 +/- 0.340</td>
</tr>
</tbody>
</table>

The RMSE of all three models are relatively low (0.244, 0.162, and 0.159), taking into consideration that the carefulness label has a range of 1.46. Models 2 and 3 have lower RMSE values than model 1. The RMSE values, which as we know share the unit of the label, and the $r^2$ values obtained, indicate good fitness of the predictive models, most especially models 2 and 3.

### 8. Conclusions

We were able to empirically validate the carefulness predictors of Shute, et al. and expand their model by adding significant predictors studied in social science research. The model derived from Shute, et al.'s features significantly predicts carefulness ($r^2 .250 +/- 0.332; \ p <0.05$) with the number of objects drawn (ObjectsDrawn) and average time before making an action on the 1st
attempt (Time1stAttempt) as the most significant predictor. From the candidate predictors’ model, carefulness has been significantly predicted ($r^2 = 0.534 +/- 0.224; p < 0.05$) by the counts of the badges (CountOfGold and CountOfSilver) and the UniqueSolutions. Consequently, this set of predictors describes carefulness as attributable to the students’ ability to solve the problems by earning gold or silver badges. If we define the badges together with the time spent as indicators of reflectivity, then, we have reason to say that reflectivity is a determinant of carefulness, i.e. reflective students tend to be careful. Further, looking at the resulting combined features model ($r^2 = 0.521 +/- 0.340; p < 0.05$), which has a better fit than Shute, et al.’s model (Model 1), we can say that the predictors of Model 2 together with the average time before making an action on the 1st attempt (Time1stAttempt) significantly predicts carefulness. This finding corroborates work that carefulness is related to mastery and reflectivity.

9. Summary of Contributions and Future Work

One contribution of this work is the empirical validation of PP developers’ model of carefulness that revealed that only the number of objects drawn and the time spent before making an action on the first attempt are significant predictors of student carefulness. Another contribution is the expansion of this model with the addition of reflectivity and mastery as significant predictors of student carefulness, confirming previous findings in social science research.

As a future work, to improve and ensure student carefulness, guidelines can be formulated on the design of educational games encouraging reflectivity among students and incorporating appropriate interventions for students with high level of mastery.

Acknowledgements

We thank the officials at the Ateneo de Davao University, University of San Carlos, University of the Cordilleras Laboratory High School and Bakakeng National High School for making this study possible. We would also like to thank Physics Playground developers and researchers, Dr. Valerie Shute, Dr. Matthew Ventura, and their colleagues at the Florida State University for collaborating with us.

References

GPT: A Tutor for Geometry Proving

Roscoe Nealle ALICBUSAN*, Laurence Nicholas FOZ*, Paolo Vittorio MERLE*, Ethel Chua Joy ONG* & Minie Rose LAPINIDb

aCollege of Computer Studies, De La Salle University, Manila, Philippines
bBr. Andrew Gonzalez FSC College of Education, De La Salle University, Manila, Philippines
*roscoe_nealle_alicbusan@dlsu.edu.ph

Abstract: Geometry is a mandatory subject for secondary school students, where they learn geometric figures and their properties, and the axioms, postulates and theorems involving them. A key topic in Geometry class is proving, where learners are required to derive and prove a certain property is true based from the given properties and by using various axioms, postulates and theorems. This is where most learners encounter difficulty. In this paper, we describe Geometry Proof Tutor, a learning environment where learners can practice Geometry proving through multiple representations - two-column proof and proof tree. With the use of a knowledge base to model Geometry concepts, the software validates the learner’s proof statements and provides corrective feedback accordingly. Test results showed that the learners found the availability of the proof tree to be useful in tracking their progress. The presence of complete reasons to use for proofs also helped them understand Geometric Proving better. However, the long list of available choices and the one-proof-statement-at-a-time user interface design made it difficult for them to encode their answers.

Keywords: Geometry, Geometry Proving, Knowledge Representation, Learning Environment

1. Introduction

Geometry is a branch of mathematics that has been around since the time of the Egyptians, when the Great Pyramids were built. It defines and relates the basic properties and measurement of line segments and angles, and Euclidean Geometry (Gantert, 2008).

Geometry is a subject taken by all students in secondary schools. Students learn geometric figures, their various properties and the axioms, postulates and theorems involving them. These lessons build upon the basics taught in elementary school - from recognizing a triangle to knowing that the sum of its angles is 180 degrees (Triangle Angle-Sum Theorem). The next and usually final step in their lessons involve proving, wherein students are required to derive and prove a certain property is true based from the given properties and the various postulates, axioms and theorems discussed in class.

Several models and frameworks have been developed to teach Geometry. The van Hiele model of Geometric Thought, which was proposed at the University of Ultrecht back in 1984, posits that there are different levels of geometric understanding which the learner must go through sequentially from simplest to the most complex level (Crowley, 1987). These five levels, arranged from simplest to the most complex, are: Visualization, Analysis, Informal Deduction, Formal Deduction, and Rigor.

Visualization encompasses the most basic understanding of geometry. It is in this level that learners recognize shapes shown to them; however, properties and characteristics of the figure are not yet evident. At the Analysis level, learners begin to examine and observe figures and would be able to determine characteristics that define them. In the Informal Deduction, learners begin to understand how characteristics and figures relate to each other. Formal Deduction, on the other hand, is the level wherein learners have the capability to see the relationship of many different factors, terms defined and undefined, as well as concepts they may or may not know, and are able to create their proofs in many ways. Finally, learners in the Rigor level can work independently from the Euclidean system and they see geometry in abstraction (Crowley 1987).

DeFT (Ainsworth, 2006) is a conceptual framework designed to make representational systems easier to understand and to maximize the benefits from multiple external representations (MERs) without too much cognitive cost for the learner. This framework proposes that the effectiveness of
MERs depends on three factors: the design factors unique to learning with MERs, the functions they serve to facilitate learning, and the cognitive tasks students must undergo when interacting with MERs.

In a Geometry class, the teacher discusses the topic with the use of diagrams. While each proof is written and discussed, learners jot down notes and simultaneously try to comprehend the lesson (Ainsworth, 2006). The teacher may then opt to give tests to assess the level of understanding of each learner, and subsequently adapts his/her approach accordingly. But checking the learner’s work requires a significant amount of time, with the teacher analyzing the students’ proofs to determine their correctness and misconceptions. This reduces the amount of time available for the teacher to give individualized feedback.

Research in intelligent tutoring systems and learning environments have designed their tutors to be adaptive to the needs of the individual learner. These tutors have been equipped with the necessary knowledge to analyze student misconceptions, to provide corrective feedback, and to devise teaching strategies to match the individual needs of the learners. In this paper, we present Geometry Proof Tutor (GPT), an intelligent learning environment for practicing Geometry proving. We discuss its strategies in analyzing errors in student proofs and providing corrective feedback.

2. Related Work

**Geometry Visualization Software:** Software has been developed to provide learners with visualizations of geometric figures or proofs. The study of Baccaglini-Frank et al. (2013) has shown positive benefits in using software to visualize geometric concepts, especially those used in proving. Geometer’s Sketchpad (http://www.dynamicgeometry.com) is a commercial software that provides its user tools to draw and create geometric shapes, objects and functions. It also allows users to alter and manipulate these figures through a simple user interface. GeoGebra (https://www.geogebra.org) is a similar free software which allows users to draw Geometric figures and also supports additional functions for Mathematics outside of Geometry at the cost of having more complex user interface. Cinderella (https://www.cinderella.de) is another software capable of visualizing geometric elements as well as other branches of Mathematics such as Calculus and Discrete Mathematics.

**Geometry Proving Software:** Tutors were also developed to facilitate learning geometric proving. Matsuda and VanLehn (2005) developed Advanced Geometry Tutor (AGT) to test two geometry proving strategies – forward chaining (proving starts at the given propositions and postulates are applied until it reaches the goal) and backward chaining (proving starts at the goal and postulates are applied in reverse until the given is reached). GeoProof (http://home.hna.org/geoproof/) is a free software which is vastly different from AGT. It does not offer the two-column proof view that most software have. Instead, it allows users to input a hypothesis regarding the given geometric figure and a conclusion; it then determines if this is true or false.

**Multiple Problem Representations:** Multiple representations are proposed by the DeFT framework as a means of assisting learning. MR Geo, a software by Wong et al. (2011), was developed to determine if displaying the geometric problem using different visualizations had any effect to students’ learning ability with regards to geometric proving. The software provided three different means of viewing the geometry problem: (1) word description, (2) static image, and (3) dynamic image. It also provided two ways of viewing the solution: (1) formal proof and (2) proof tree.

**Feedback System:** Feedback systems are essential for tutoring systems as it is the means for them to help their users in learning. Fluckiger et al. (2010) cites that learning is a cyclical process of continuous self-assessment and therefore, feedback generally facilitates this. Nyquist (2003) classifies this assessment into categories, namely: (1) weaker feedback only, (2) feedback only, (3) weak formative assessment, (4) moderate formative assessment, and (5) strong formative assessment. Previously mentioned software fall into the first category, which only inform users if their inputs are correct or incorrect and nothing further. Similarly, the DOST Courseware Project (http://courseware.dost.gov.ph) is a tutoring system designed by the Philippine's Department of Science and Technology for secondary level Science and elementary and secondary level Mathematics. It stops the user from proceeding further upon entering incorrect answers; however, nothing else is mentioned. SalinLahi III, developed by Regalado et al. (2015), is another tutoring software designed to tutor heritage language learners in Filipino grammar and sentence construction. SalinLahi III gives more
comprehensive feedback using template-based NLG by categorizing the user input as (1) correct, (2) incorrect, (3) near the answer, and (4) cannot be understood.

3. The Geometry Proof Tutor

GPT maintains a bank of Geometry proving problem sets that are designed by teachers through the Editor facility. Every proving problem has a set of similar characteristics (Bass et al., 2001). These are the diagram visualizing the geometric figure, the written statement of the given geometric figure along with the statement to be proven, the statement of facts considered to be true, and the reason for validation (why the statement is true).

3.1 Presenting Geometry Proving Problems through Multiple Representations

Geometry proving problem sets are answered through the interface shown in Figure 1. As Duval (1998) pointed out, learning Geometry involves three cognitive tasks: visualization, construction and reasoning. When deriving a Geometry proof, the learner must be able to understand the given problem from the problem statement, and the figure that comes along with it. The learner then uses his/her knowledge in Geometry and proving to derive proof statements that may lead to the conclusion.

There are three types of representations that are useful for geometric proofs (Wong et al., 2011) - the problem representation, the visual representation and the proving representation. As shown in Figure 1, the problem representation (found on the left side of the user interface) is in the form of a text to help the learners understand mathematical symbols and language, and the logical relationship between all given conditions and the end goal condition. The problem representation is derived from the test bank of proving problems as defined by Geometry teachers.

The visual representation complements the problem representation, to allow the learner to infer from one representation the information that may be hidden or lacking in the other. This approach can help strengthen their understanding of the problem as well as help them find clues on how to get to the solution (Wong et al., 2011).

![Figure 1](image_url) Space for Practicing Geometric Proving

The third representation contains the proof made by the learner. There are two types that can be used: the traditional two-column proof, and the proof tree. Figure 2 shows a sample of the two-column proof, consisting of statements and their corresponding reasons. The two-column proof is shown in the middle part of the user interface (Figure 1). The proof tree is a tree-like structure to represent the network of inferences (Matsuda & VanLehn, 2005). It has the following properties (Wong et al., 2011):
3.2 Proof Checking and Returning Feedback

Whenever the learner inputs a statement to be added to the proof, GPT evaluates the statement through a series of steps. To better illustrate this, the proving problem shown in Figure 2 will be used as a running example.

The first step is a simple preliminary check to look for special cases, such as when the learner identified the statement as part of the given. If so, the GPT checker must cross-check the statement with the list of given. For example, the first statement added to the proof in Figure 2 has “Given” as its rationale. As such, in this step, this statement is cross-checked (Line AB is perpendicular to Line BC) in the list of given. The special case is vastly different from any other cases and as such, would diverge completely from the other steps. That is, if a statement is found to be a special case, all succeeding steps would be skipped.

The second step involves template checking, which simply verifies the template used by the statement and if there is a rule which results in that statement. For instance, for the second statement in Figure 2, GPT checks if there exists a rule which results in the template “[A1] is a right angle” and would find the “Right Angle Theorem” rule.

The third step checks if the stated rationale is correct and there exists a subrule corresponding to it. Subrules correspond to the different rationales under the rule. For instance, under the rule “Triangle Congruence”, the subrules include “Side-Side-Side (SSS) Congruence Postulate” and “Side-Angle-Side (SAS) Congruence Postulate”. For some rules where there only exists one subrule (meaning that there is only one postulate, axiom or theorem that would result in that statement), their names would be similar. In the running example in Figure 2, this would be the case - “Right Angle Theorem” rule and subrule.

In Geometry proving, it is common for one proof statement to rely or be related to previous proof statements. Thus, the final step involves checking the referenced statements used in the proof. For instance, the “Side-Side-Side (SSS) Congruence Postulate” states that two triangles are congruent if their sides are congruent with one another. As such, entering this rationale into GPT would require three proof statements corresponding to each pair of sides which are congruent. In the running example in Figure 2, the second statement would require a statement stating that the two lines are perpendicular to one another. As such, the previous statement (“Line AB is perpendicular to Line CD”) would be referenced for the newly added statement to be valid.

If at any point, the GPT checker finds an error in the student’s proof, the checking stops and the incorrect proof statement is highlighted with a corresponding feedback as a form of weak formative assessment. According to Nyquist (2003), weak formative assessment gives feedback coupled with some explanations. As such, the generation of feedback also utilizes templates. The templates used are divided into two categories - templates for correct answers (e.g. “You are correct!”) and templates for
wrong answers (e.g. “The statement [A1] is not reflexive. The reflexive property will only result in statements of congruency.”).

The predefined templates are not commutative. There are instances wherein some of the referenced geometric figures can be interchanged. An example of this is the congruence template, where if Figure $A$ is congruent to Figure $B$, then it should follow that interchanging the indices in the template, i.e., Figure $B$ is congruent to Figure $A$, would be equivalent. However, the non-commutative design of the templates does not support this. That is, GPT does not have the knowledge that some indices can be interchanged. This can be problematic when learners try to interchange the indices with the idea that they are equivalent. It also introduces problems during checking. For example, in the SSS Congruence Postulate checking, if the learner’s input goal is to prove that Triangle LOV is congruent with Triangle LEV, the checking algorithm will return errors should the learner interchange the congruence statements (i.e. VE is congruent with LO instead of the other way around).

4. Evaluation, Results, and Discussion

GPT was evaluated by 23 tertiary level students. Each participant used GPT to solve a proving problem. The participant’s proofs and GPT’s responses were internally logged for further analysis. A debriefing interview was conducted to solicit feedback regarding their experience. They also accomplished survey forms to provide a quantitative evaluation of the software.

Table 1 summarizes the average evaluation scores from end-user surveys, using a rating scale of 1 to 5, with 1 representing strongly disagree, 2 for disagree, 3 for neutral, 4 for agree, and 5 for strongly agree. Overall, GPT received an evaluation score of 4.08. The first criterion, “I can easily add proof statements.” got the lowest score. This is attributed mainly to the difficulty encountered by the participants in searching for reasons due to the presence of too many choices which are not organized in any manner. Furthermore, proof statements can only be entered one at a time.

Table 1: Evaluation Scores of GPT from the First Iteration of End-User Testing

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can easily add proof statements.</td>
<td>3.91</td>
</tr>
<tr>
<td>The representation of the problem is appropriate.</td>
<td>4.09</td>
</tr>
<tr>
<td>The proof tree assists in showing me how my proof has progressed so far.</td>
<td>4.22</td>
</tr>
<tr>
<td>The proof tree is easy to read and understand.</td>
<td>4.17</td>
</tr>
<tr>
<td>The feedback was helpful in letting me know why my proof was wrong.</td>
<td>4.00</td>
</tr>
</tbody>
</table>

The rest of the criterion received scores above 4.00, with participants stating that they can easily see the possible answers. They found the feedback to be helpful in identifying errors in their proofs. The presence of complete reasons to use for proofs also helped them understand Geometric proving better. Among the criterion, the proof tree received the highest score. It was observed that participants interacted with the proof tree, with some of them moving the view around to see the parts of the tree obscured from their vision. They also used the proof tree as a guide to determine how far their proof has progressed and what their next step would be.

5. Conclusion and Future Work

Geometry Proof Tutor is a learning environment that evaluates the proofs and inferences placed as input by the learners, and gives corrective feedback accordingly. The use of multiple representations allows learners to input their answers using the two-column proof format, and view the corresponding visual
representation generated by the software as a proof tree. An editor facility is also included to allow
teachers to define Geometric proving sets.

Test results showed that the participants found the proof tree and the feedback to be useful
during a learning session, as these enabled them to track the progress of their proofs and to understand
the reasons behind incorrect answers. However, the user interface design needs to be improved to
enable a more efficient mechanism for the learners to input their proof statements, including the lookup
facility to search for reasons.

Currently, the logs that note down why a learner’s proof statement is valid or not (Figure 2), and
the retention of the student’s mistakes are used to evaluate the knowledge of the learner in geometric
proving. However, this only holds for the current problem set. Since GPT is not a full-blown ITS, it
does not maintain a student model to track an individual learner’s understanding of the different topics
in Geometry. Mistakes committed by a student in one session are not stored and used in subsequent
sessions. Future research should consider the use of a student model in determining a specific learning
goal for each session in GPT. The generation of feedback can also be tailored to the performance and
needs of the individual learner.

The algorithm that checks a learner’s proof relies heavily on the knowledge representation of
Geometry properties, axioms and postulates. As such, it is unable to correctly check a proof if the said
proof is not represented in GPT. Examples of these are rules such as conic sections or arcs, which are
currently not included in GPT’s scope. Other examples include special proving methods outside the
traditional proving methods, such as proof by construction - a proving method wherein an additional
figure is added to the given figure to give light to more properties.

Although the use of templates to represent Geometry concepts currently has its limitations,
amassing learners’ usage of these over a period and across different learners can generate data for
educational analytics. Future work can conduct an analysis of these data to find patterns on how learners
use and misuse axioms and theorems in coming up with their proofs. Similarly, data captured in the
working memory that GPT uses when checking a learner’s proof should also be stored and mined.
Insights on common misconceptions can then be gleaned from these two data sources and subsequently
be used by teachers in designing teaching materials to address difficulty that learners encounter in
Geometry proving.

References

Learning and instruction, 16(3), 183-198.

geometry. PNA, 7(2), 63-73.

Changing World. Prentice Hall.


Crowley, M. L. (1987). The van Hiele model of the development of geometric thought. Learning and teaching
geometry, K-12, 1-16.


partners in assessment to enhance learning. College teaching, 58(4), 136-140.


Nyquist, J. B. (2003). The benefits of reconstruing feedback as a larger system of formative assessment: A


Inquiry-based Support System to Improve Intention Sharing Skills

Natsumi MORI*, Yuki HAYASHI & Kazuhisa SETA
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan
* mori@ksm.kis.osakafu-u.ac.jp

Abstract: In order to achieve fruitful and creative discussions, it is important that the speakers verbalize and share their own intentions behind the utterance properly with the listeners. We call such essential skills for creative discussion “intention-sharing skills.” The research objective is to propose a research activity support system for cultivating novice researchers’ intention-sharing skills. To achieve this objective, we focus on the opportunities of research meeting. We, especially, take notice the researchers’ internal self-conversation processes of organizing their own thoughts to improve their readiness for sharing intentions in a following research meeting. In this paper, we first structure inquiries for phase of organizing own thoughts. Then, we develop a research activity support system in which researchers focus on organizing the structure of their own thinking processes as a pyramidal structure consists of chains of inquiries and answers.

Keywords: Intention sharing skills, structured inquiries, inquiry-based support

1. Introduction

In a conversation, participants try to grasp not only superficial meaning of the speakers’ utterance but also their implicit and tacit intentions behind their talks with the clue of the listeners’ pre-existing knowledge, the speaker's nonverbal behaviors and the underlying contexts. To share intentions of each utterance is the key to a successful communication among participants (Tomasello et al., 2005). To make a conscious intention sharing is essentially required in the knowledge co-creative discussion: speakers themselves should try to clarify not only factual and episodic information but also their intentions behind the utterance (Enrici et al., 2011; Velleman, 1997). In this research, we call such tacit but essential skills of speakers as “intention-sharing skills” as a kind of metacognitive skills. If speakers cannot well display their intention-sharing skills, the discussion sometimes becomes confusing because of a discrepancy between the speakers’ intentions and the listeners’ understanding.

In this research, we focus on “research meetings.” In a research meeting, discussion subjects are sometimes on ambiguous and unclear conceptual ones and only the presenters themselves know their own research progress from the last meeting. As mentioned above, sharing contexts of one’s internal self-conversation processes with others is an essential activity for creative discussions. In order for presenters to explicitly share the contexts with other research collaborators, they need to understand their own thought processes comprehensively in advance, and display their intention-sharing skills expressing their thought processes. Since the research meetings have been carried out continuously to improve the research, it is reasonable and best-case opportunities whereby one practically cultivates own intention-sharing skills that require for a long-term period.

In this paper, we propose a research activity support system for cultivating learners’ intention-sharing skills. We focus on learners’ internal self-conversation processes to organize their own thoughts and intentions before a research meeting in order to improve their readiness for verbalization to share their intentions in the following meeting. First, we propose a learning activity model to cultivate intention-sharing skills, and discuss the concept of structured inquiries which prompt their internal self-conversations for organizing their thoughts. And we develop a research activity support system and show its functions.
2. Approach

2.1 Learning Model of Intention-Sharing Skills

Sharing skills of one’s intentions with others play an important role to achieve fruitful and creative discussions. To cultivate such skills, it is important to conduct continuous performance of one’s internal self-conversations which improves their readiness for sharing intentions in the following discussion and reflection of their discussion processes with referring to their internal self-conversations.

Figure 1 shows our learning model for cultivating intention-sharing skills through research activities. This model is composed of four phases: (a) organizing own thoughts, (b) systematization of thoughts, (c) sharing of intentions and (d) modification of thoughts. In the following, we describe the activities of learners performed in each phase. Intuitively, the first two phases (a) and (b) correspond to performing internal self-conversation processes, (c) does displaying intention-sharing skills at a research meeting and (d) does reflection of (a), (b) and (c).

- **Organizing own thoughts (Fig. 1(a))**: The phase at which learners try to clarify their unconscious thinking processes through activating their internal self-conversation before a research meeting. The thinking activity is essentially invisible and tacit processes. In the phase, learners explicitly express their thoughts as tree structure (Minto, 2006) which consists of the chain of inquiries and answers in a top-down fashion. Here, learners also need to express the relationships between respective structures of their research and the general structure of the target research domain.

- **Systematization of thoughts (Fig. 1(b))**: The phase at which learners try to specify purposes of the research meeting based on the organized thinking process in phase (a). In order to improve their readiness for sharing intentions in the following discussion, it is necessary to deeply think and choose the information should be shared among colleagues. Then, the documents need to be systematized which clearly describe the objectives in a logical manner.

- **Sharing of intentions (Fig. 1(c))**: The phase at which learners try to share their intentions at the meeting. It is required that they should not merely express the facts but verbalize their intentions.

- **Modification of thoughts (Fig. 1(d))**: The phase at which learners reflect on the discussion by modifying the tree structure of their thoughts created before the meeting. They are ready to criticize their thoughts objectively if elaborate internal self-conversations are performed in phases (a) and (b). The colleagues’ opinions and questions in the discussion give an ideal opportunity to find a lack of their own thoughts in terms of intention sharing and internalize them for the next meeting.

As described above, our model is designed to cultivate learners’ intention-sharing skills step-by-step through the cyclic research activities.

In this paper, we focus on the phase of organizing own thoughts (Fig. 1(a)) out of the above four phases. In order to improve their readiness for sharing intentions in the following discussion, it is essentially important to organize their thinking deeply in their internal self-conversations in phase (a).

![Figure 1. Learning Model for Intention-Sharing Skills](image)

As a representation method of organize thoughts, we focus on a tool called mind map (MM) which has a predilection for the pyramid principle (Minto, 2006). The MM is a tool that adapts well to the structure of semantic memory of the human brain. By grace of making the best use of the structure, the tool contributes to organizing, understanding and memorizing information faster (Davies, 2011). While MM is designed for domain-independent general-purpose tool. But even if the target topic is learners’ own research contents, it is difficult for learners to verbalize their intentions behind their thoughts into a pyramidal structure. In many cases, as Minto pointed out, this problem is caused by the reason that even writers themselves cannot be conscious of what they think without externalization.
Therefore, in order to prompt organizing learners’ thoughts along with their thought contexts, we conceive the idea of providing inquiries (Ash & Clayton, 2009) which reflect the structure of target research domains as a stimulus. In the following section 2.2, we describe inquiries that should be considered in the phase of organizing thoughts (Fig. 1(a)) and structure a model of the inquiries.

2.2 Structuring Inquiries for the Phase of Organizing Own Thoughts

In the phase of organizing own thoughts (Fig. 1(a)), it is important for learners to overlook and situate their own thoughts about research so as to deeply understand their own research. This activity contributes to changing biased thoughts to fair and balanced ones. In accordance with the balanced thoughts, they can understand their thinking processes such as why I need to think about this, what is unclear, what matter I need to think about first, and why I should discuss it in the meeting, and they get ready for specifying the purposes of the research meeting. Followings are five necessary thinking activities that correspond to the policies of specifying inquiries to stimulate learners’ internal self-conversations for their own research.

A. Consideration both from macro- and micro-viewpoints: In order for learners to delve into their own research deeply to verbalize their intentions, it is important that they should consider both the domain-independent cross-sectional research structures and the domain-specific structures in a balanced manner. To constantly consider their own research from these macro- and micro-viewpoints contributes to display well-balanced and critical thinking skills to improve their research and also to sharing their intentions among participants.

B. Understanding of research structures: Through the activity of verbalizing and structuring their own research, learners are aware of the irrational points of their own thoughts such as logical contradiction. It contributes to clarifying the purposes of the discussion and the points of sharing their intentions.

C. Setting of criteria for decision-making: In order to achieve a consensus-based decision-making among research collaborators, it is essentially required to verbalize the criteria of each decision-making. This activity requires learners to grasp and verbalize the logical paths of own thoughts leading to the decision-making.

D. Clarification of reasons and purposes: This activity requires learners to verbalize not only the results of thoughts (contents of leaf nodes in the pyramidal structure) but also the causal and logical paths to them by considering their reasons and purposes.

E. Critiquing rationality: In order for learners to organize their thoughts in a rational manner, it is also important to critique the rationalities of each verbalized thought. For promoting above thinking activities, we organized a set of both domain dependent and independent inquiries to stimulate them in their internal self-conversations. The inquiries prepared for promoting above activities A, B and C specify the requisite inquiries to promote research activities. They are organized to provide stimuli along their thought contexts. On the other hand, the inquiries prepared for the activities D and E play the role of reconsidering verbalized thoughts of the activities A, B and C.

It is desirable to systematize these inquiries in advance in terms of the commonality/differences among them in a context of a research domain rather than just listing them in a chaotic fashion. Therefore, we organize the inquiries based on the ontological engineering approach. Figures 2, 3 and 4 show a part of the systematized inquiries, each of them conceptualized along the “is-a,” “part-of” and “attribute-of” relationships to provide the inquiries to prompt learners’ thinking activities A, B and C, respectively. In these figures, the nodes marked red flags indicate the concepts and the unmarked ones correspond to the inquiries. As shown in Fig. 2(i), the inquiries about research (e.g., “What is the objective?”) are connected with the concepts (e.g., “Objectives”). This systematization allows us to easily maintain the set of same/different meaning of inquiries under the concepts and thus easily grasp the structure of inquiries.

For activity A, it is desirable that learners constantly consider their research from both viewpoints of abstract and concrete, in other words, from domain-independent cross-sectional and domain-specific ones. We use “is-a” relationship to organize the concepts to represent these macro- and micro-viewpoints. The concept “Problems,” for instance, has a child concept “Difficulties for learners” that should be considered in the research field of intelligent tutoring systems (Fig. 2(ii)).

In order to promote activity B, we systematize the concepts about research task structures based on “part-of” relationship as shown in Fig. 3; it is necessary for learners to think about “Problem-finding
processes,” “Task-setting processes” and “Task-solving processes” in “Problem-solving processes.” Fig. 3(iii) shows the relations of these concepts. By organizing this way, the system could provide the inquiries that should be considered in their research task structures.

In activity C, it is necessary for learners to view the properties of a concept as the criteria for decision-making. In order to represent the concept properties, we apply “attribute-of” relationship. Figure 4 represents a part of the systematized concepts. The concepts, for instance, “Methods for problem-solving” and “Teaching materials,” each of them requires to consider their properties to conduct educational system research such as “Characteristics” and “Problems,” are defined by “attribute-of” relationship by referring to (Mizoguchi, 1995).

In addition, we defined the inquiries such as “Why do you think so?,” “What is the objective?” for the activities D in order for learners to consider the reasons and purposes related to all the concepts. For the activities E, we currently defined an inquiry “Why do you think these are rational?” to examine the rationalities of relationships among each verbalized thought.

In this research, we embed the inquiries into a research activity support system in order for the system to intellectually provide useful inquiries by capturing learners’ thought (see Section 3).

3. Inquiry-based Support System for Cultivating Intention-Sharing Skills

We have already confirmed that even simplified version of inquiries prompts learners’ internal self-conversations and contributes to verbalizing their intentions (Mori, 2016). Based on the experimental results, we specified a structure of inquiries from the viewpoint of intention sharing as discussed in section 2.2, and developed a support system for learners to organize their own thoughts about own research using structured inquiries.

Figure 5 shows an interface of our system. In the area of thought expression map (TEM) (Fig. 5(I)), learners can organize their thoughts as sequences of chosen inquiries (blue-colored node) and respective their answers (orange-colored node) based on the concept of pyramid principle (Minto,
In this area, thinking processes of learners are verbalized in a step-by-step manner by repeatedly verbalizing the chains of inquiries and answers in their internal self-conversations. The inquiry listing area (Fig. 5(II)) displays the systemized inquiries in a hierarchical list format. Learners can spontaneously choose the inquiries to construct their own thought contexts. While paying regards to the learner’s own thinking, the displayed inquiries play a role of stimuli as a scaffolding along their thought contexts. In addition, based on the learners’ selected inquiries from the listing area, the system detects an upper concept of the inquiries along ‘is-a’ relationship and displays the related inquiries of the concept based on the relationships among concepts (Fig. 3 and 4). In this way, by adaptively providing the inquiries related to the learners’ thought contexts on the fly, it aims at activating their further internal self-conversations and improving the quality of the thought verbalization activity.

The inquiry listing area consists of three kinds of sub-areas for promoting five kinds of activities discussed in section 2.2. Each area is further divided from the viewpoint of timings to provide respective inquiries for learners. The inquiries classified in each area and their timings of display are explained below.

- **Inquiry area for verbalizing own thoughts (Fig. 5(i)):** In this area, the inquiries for the thinking activities A, B and C described in section 2.2 are shown. In the following, we describe how to provide the inquiries for prompting each thinking activity. To answer them makes learners being conscious of their intentions.
  - **Activity A:** In order to enhance learners’ fair and balanced thinking from both macro- and micro-viewpoints, the inquiries, which correspond to the instances of macro-/micro-viewpoints concepts, are hierarchically provided along the “is-a” relationship. For example, since the micro-level concept “Difficulties for learner” is connected with macro-level concept “Problems” via “is-a” relationship (Fig. 2(ii)), its corresponding inquiry “What is the difficulty for learner? (micro-level)” appears under the inquiry “What is the problem? (macro-level)” as shown the area surrounded by a black border in Fig. 5(i’).
  - **Activity B:** For prompting learners’ understanding of the research domain structures, the provided inquiries are dynamically narrowed down according to the selected inquiries by learners. Figure 5(i’) shows an example in which a learner selects the inquiry of the concept “Experiments.” In this case, the inquiries of the concepts that correspond to its parts (e.g., “Experimental objectives”) or sibling relationships (e.g., “Evaluations”) defined as “part-of” relationships (Fig. 3).
  - **Activity C:** In order to make learners be aware of the logical structures of own thoughts leading to a decision, the inquiries shown vary with the selected inquiries. The inquiries are narrowed down based on the concepts which have “attribute-of” relationships as shown in Fig. 4. When a learner selects the inquiry of the concept “Methods to approach,” for instance, the inquiries about properties (e.g., “What are its characteristics?” and “What are its problems?”) are appeared in inquiry listing area. These inquiries play a role of stimuli for rational decision making in a situation of having multiple options and to answer them makes learners being conscious of their own intentions of their decision making.
• **Inquiry area for verbalizing intentions (Fig. 5(ii))**: In this area, the inquiries for promoting the thinking activity $D$ are displayed. The inquiries include “Why do you think so?” and “What is the objective of it?” to prompt verbalization of reasons and objectives of one’s thought. Since these inquiries are essentially important and thus the learners should keep in mind at all the times, the inquiries are always shown statically in this area regardless of the learners’ thoughts on the TEM.

• **Inquiry area for considering rationalities (Fig. 5(iii))**: In this area, the inquiries for the thinking activity $E$ are displayed. For example, when a learner selected one of two inquiries of the concept types “experimental objectives” and “experimental procedures,” their answers expressed in the TEM must hold a rational relationship, the frames of these two nodes (inquiries) on the TEM are highlighted with green color so as to let learners be aware of considering their rationalities. Learners can think the rationality by choosing green node(s) that they want to focus on. In the TEM, the selected nodes are highlighted with pink color as shown in Fig. 5(III) and the inquiry “Why do you think they are rational?” appears in this area.

The classification of inquiries into these areas aims to encourage learners’ internal self-conversations from domain-level thinking activities (A, B and C) to meta-level thinking activities (D and E).

4. Concluding Remarks

In this research, we proposed a research activity support system for cultivating novice researchers’ intention-sharing skills through their daily research activities in a laboratory. We focus attention on the activity of organizing own thoughts in the internal self-conversation processes in order to improve their readiness for sharing intentions in the following discussion. The proposed system allows learners to organize the structure of their thinking processes as a pyramidal structure consists of chain of inquiries and answers. The system also provides inquiries which play a role of stimuli for learners’ comprehensive thinking activities. Since the inquiries are structured on the basis of the ontological engineering approach, the inquiries capturing their thought contexts on the TEM are dynamically provided.

Currently, we installed our support system into the learners’ daily practical research activity. Still in a preliminary stage, we get a sense of possibility of sustainable use of the system in daily actual research activities in a laboratory. For future work, we improve the usability of the system and conduct a detailed analysis of learners’ growing process of intention-sharing skills.

References


A Tool for Data Acquisition of Thinking Processes through Writing

Wasan Na Chai*, Taneth Ruangrajitpakorn & Thepchai Supnithi
Language and Semantic Technology Laboratory,
National Electronics and Computer Technology Center, Pathumthanee, Thailand
*wasan.na_chai@nectec.or.th

Abstract: Unclear thinking is a great obstruction in outward communication, especially writing. In this work, we present a tool to support in becoming aware of one's own thinking via writing. The tool is designed to slow down a writing task and spend more time in realising thinking by clarifying contents. Users are asked to express concepts of each clause and connection of clause to another. By letting learners to slow down and think more clearly about what to write, we expect that their thought will become clearer and thus will produce a better output. On usage, the tool records actions made in the tool and changes in revision. These details can assist a supervisor to realise the root of confusing thought and comment accordingly and pertinently. From an experiment, a written output made via the tool gained a higher quality comparing to an output written in traditional environment for each individual sample. We also found that the top rated written outputs via the tool were made after being clear on content declared with the tool requirement. This can be implied that awareness about thinking can help in improving their writing result. Last, samples revealed that using the tool can help them practising in becoming aware of thinking.

Keywords: Metacognition, writing skill, skill learning, awareness of though, collaborative tool

1. Introduction

Thinking is the most crucial process that human automatically and implicitly does in all activities. It is a core of our actions such as writing, creating, planning, etc. A thought produced from careful thinking processes such as a logical statement, riskless plan and design is considered to be better than a rush idea without thinking. However, thinking is automatic cognitive process and requires a practice and experience to get improved. One of common methods to improve thinking is to realise one's own thinking within thinking processes. Thinking about thinking (metacognition) [Flavell, 1979] is one of higher-order thinking skills that can be used to clarify cloudy thoughts in thinking processes. Writing is a solid evident of one's thinking [Brown, 1987] [Wong, 2005] since a written output such as an essay is a traceable evident that can reflect person's thinking via a choice of words, style, and logical connection of sentences. A written piece lacking readability from poor discourse (unconnected statement or without direction in convincing) can be implied as confusion in thinking processes from its writer. This issue can often be seen from novice writers who do not excel in thinking or planning their thoughts while written outputs of veteran writers rarely display with this issue. Hence, the more the thinking is clear and connected; the better of the content can be expected.

In this paper, we aim to assist in clarifying one's own thinking in cognitive processes. We hypothesise that the thinking can be much clearer if a person becomes more aware of own thinking and clearly realise in content of a though. We thus design a tool to help on declaration on a content produced while writing to help increasing awareness of thinking. Hence, users can practice to become aware of thinking and its content through writing. Similarly to our idea in concerning in thinking processes, there is a tool called Swan - Scientific Writing AssistaNt [Kinnunen et al., 2012] aiming to assist a user in writing focussing on fluidity of concepts. Its specification is to guide with the content, not the grammar.
or spelling. The tool provided a predefined list of necessary contents for each section as a predefined structure of contents. This tool also includes a function for evaluation to detect fluidity of contents. Fluidity represents how well a text flows from sentence to sentence in which helps to connect content in unison. Though the tool promotes importance in connection of sentences, it does not practically aim on practicing in a cognitive process of becoming aware of thinking for self-clarifying in content for writing.

2. Design of the tool

The aim of this work is for students who lack awareness of their thought/plan while writing. The main focus is to scope their content based on the thought/plan. Thus, the system is designed as a supporting system to clarify their thought and collaborate with writing expert as a coach to guide through their expression. The system consists of three main functions as 1. Obtaining writing input from a learner, 2. Recording learner’s behaviour in writing and tracking their progress, and 3. Providing a reaction or a comment from a coach team. A layered architecture of the proposed system is sketched in Figure 1.

![Figure 1](image)

**Figure 1.** A layered architecture of the proposed training ground for metacognitive writing skill.

The tool is separated into three parts. The first part is a user-interface designed to get a writing input along with its information and to communicate among learners and coaches. Secondly, The process is for detecting learners’ behaviours and managing writing inputs. Last, the storage is designed to record the writing and comments with traceable records, and the records can be generated into a log for reference and recall. The tool is designed to handle a writing piece as project based; thus, a learner can focus into their writing task. However, the system allows users to access contents in their authorised projects as reference and example.

2.1 Design of User-interface

Since there are couple of roles involved in the usage and different in their task, we separately design each user-interface (UI) for each role. User roles in this work include learner, coach and admin. The first and foremost role in this work is learner. Learners are to use the system as a ground to provide their writing and its information. The key of comprehensibility is to realise the thought while writing. We therefore design to get an input in smallest context based on action (verb) as a clause. At first, learners have to select an overview of a content regarding to standard sections of publication such as *introduction, methodology and experiment*. The input of learners consists of information mentioned in Table 1.
Table 1: Learners’ input and its details

<table>
<thead>
<tr>
<th>Input type</th>
<th>Description</th>
<th>Possible value/ Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Writing in a clause level</td>
<td>Free text in a clause level: ‘you can see many of write with no metacognitive skill (an example from actual writing input that may not be grammatically and rationally correct)</td>
</tr>
<tr>
<td>Type</td>
<td>Type of intention selected from a predefined list (one choice)</td>
<td>• Fact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o General Fact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Specific Fact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Detail Fact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Providing Definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Example</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Proposed Statement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Opinion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Evidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reason</td>
</tr>
<tr>
<td>Key Concept</td>
<td>Summary of concept</td>
<td>Short text to represent the core concept of the clause.</td>
</tr>
<tr>
<td>Related</td>
<td>Indication of relation to another clause</td>
<td>Numeral index pointing out a relation of a current clause to an existing clause – only need for some types such as example, reason and condition</td>
</tr>
<tr>
<td>Free text Comment</td>
<td>Text to remark or communicate to coach</td>
<td>Additional free text to convey note to coach or as a remark</td>
</tr>
</tbody>
</table>

2.2 Behaviour Detection and Log Generation

To monitor writing for metacognition, writing behaviour is a key to determine learners’ thinking process. We design the tool to record not only writing content but also an interaction of learners in giving a writing input. For a learner side, the detection of interaction includes a timestamp of actions and an order of given input types. These data can inform how long the content have been created, user’s’ thought before, after, in between writing and an overview of behaviour. By combining these data to a quality of writing content judged by coach, we can understand learners’ psychological process in writing and infer a type of behaviours. For example, if a learner had written the clause for 15 seconds but spent time in selecting a content type for more than 30 seconds after finishing writing and the content is obviously confusing, it can be roughly inferred as ‘the learner did not think before writing and was not aware of what to express’. Therefore, the tool is designed to record details given in Table 2 as behaviour data.

Table 2: Data of interactions from learners in using tools

<table>
<thead>
<tr>
<th>Detection type</th>
<th>Interacting with Input type</th>
<th>Recoding information</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timestamp</td>
<td>Typing in Writing panel</td>
<td>• Start time</td>
<td>Calculation for a duration of writing content</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• End time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clicking Dropdown of content type</td>
<td>• Start time</td>
<td>Calculation for a duration of selecting a content type from the list</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• End time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Typing in Key concept panel</td>
<td>• End time</td>
<td>Recording time a key concept is provided</td>
</tr>
<tr>
<td>Order of Action</td>
<td>Writing panel, Dropdown of content type and Key concept panel</td>
<td>• Order of action</td>
<td>Recording order of actions based on time and duration</td>
</tr>
</tbody>
</table>

82
The captured behaviour data are to record for each of content given via the tool. They are recorded along with the given clause for both new clause and edition. The content and behaviour are recorded in a database, and they will generate to two types of a log detail, i.e. a content log and a full log. A content log is a history of only writing for each clause. A full log includes both content and behaviour details. On a coach side, it is possible for a writing project to have two or more coaches. All given comments and decision of passing status from all involved coaches are recorded with a timestamp for managing consequence into a comment log stored in a database. For a learner, they are able to view details in both a content log to track a change in their writing and a comment log of the current project. Learners are also allowed to access a content log of his/her past project(s). For coach side, involved coaches are authorised to access on a full log to monitor writing result and behaviour to determine learner’s writing thinking process. Moreover, a coach can view a comment log to observe comments from other coaches.

3. Evaluation and Discussion

For evaluating usefulness of the proposed tool, an experiment in usage was set up. The main focus is to compare how learners perform in writing within and without an environment that they are forced to slow down their thinking about what to write via the tool. Moreover, we also want to study how effective in writing by pointing out the misconception in thinking via a strict environment to explicitly mention their idea/thought (from the tool) can bring.

The participants were ten volunteer Thai graduate student and two experienced researchers. The students had completed their thesis topic examination. The chosen experienced researchers had published over 15 academic publications and excelled in guiding for paper writing. Learners were asked to write two contents: 1) an abstract of their thesis topic via the tool and 2) a short summary of their related work via a normal word processor (without the tool). The writing pieces must be in proper English in technical language and was limited to 20 clauses in total. All learner samples past English standard qualification from their respective university and have an average of TOEFL score in PBT about 450 points. In this experiment, revision of each clause was limited to 3 times at most since we did not intend to see how fast learners can improve, but how the tool affects learners. The samples were randomly split into two groups. The first group is to use the tool for writing abstract first and write a short summary later in one-week interval while the second group did in vice versa.

Coaches were instructed to give only guidance within a scope of idea/thought. We observed the usage with two observers and recorded their actions through a log generated by the tool. Learners were allowed to access to previously made thesis reports for referencing. There was no time limit for completing the task. For coaches, each clause in every revision was rated for two aspects as follows. 1) Matching of content and given information (key concept and content type) and 2) Properly connecting idea/thought in logical relation leading to the objective.

The scale was from 1 (worst) to 5 (best). For the first aspect from the writing without the tool, a key concept and content type were asked later after finishing the experiment for rating. Figure 2 shows average rates for each learner from the first aspect and second aspect, respectively. The bars of each learner represent an average rate of all the initial clauses (before revision from obtained comments) in comparison to a writing piece made via the tool and via common word processor.

From the results in Table 2 and 3, we found that the rates of the two writing piece made via the tool were higher. Based on samples’ ability, we can see that learner id#4 and id#9 were above average since their rates were the tops, and their writing pieces were over moderate in both setup environments. However, they eventually gained higher rates for the piece they made on our tool, especially for the aspect of properly connecting though to the objective. We also noticed that low proficient learners such as id#3 and id#10 gained a significantly increasing rate for the aspect of matching content to their thought. In terms of statistics of writing result and process, the notable details are summarised into Table 3. We found that learners spent more time when using the tool to generate a clause. With the intention for a writer to slow down process to collect their thought, we found the usage results compromised with the aim. By combining results from Table 3 and Rated results, we noticed that a time spent in thinking about content type and key concept individually affected the rated results except for learner#4. After interviewing, we realised that the id#4 was different than other and can be recognised
as an outlier who had experienced in writing several publications beforehand while other learners are fairly new to the task.

![Figure 2. A comparison of rated results regarding from coaches.](image)

**Table 3: Statistical details of writing results**

<table>
<thead>
<tr>
<th>Learner id</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract (a piece with tool)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clause amount</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Total time (s)</td>
<td>620</td>
<td>688</td>
<td>491</td>
<td>489</td>
<td>593</td>
<td>611</td>
<td>583</td>
<td>771</td>
<td>654</td>
<td>466</td>
</tr>
<tr>
<td>Average time per clause</td>
<td>88.57</td>
<td>86.00</td>
<td>61.38</td>
<td>48.90</td>
<td>98.83</td>
<td>87.29</td>
<td>64.78</td>
<td>128.50</td>
<td>81.75</td>
<td>77.67</td>
</tr>
<tr>
<td>Median time per clause</td>
<td>68</td>
<td>69.5</td>
<td>54.5</td>
<td>44.5</td>
<td>95.5</td>
<td>88</td>
<td>60</td>
<td>101.5</td>
<td>77.5</td>
<td>72.5</td>
</tr>
<tr>
<td><strong>Short summary of related work (a piece without tool)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clause amount</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Total time spent (s)</td>
<td>572</td>
<td>480</td>
<td>354</td>
<td>498</td>
<td>521</td>
<td>424</td>
<td>512</td>
<td>388</td>
<td>410</td>
<td>373</td>
</tr>
<tr>
<td>Average time per clause</td>
<td>71.50</td>
<td>53.33</td>
<td>44.25</td>
<td>45.27</td>
<td>74.43</td>
<td>60.57</td>
<td>51.20</td>
<td>55.43</td>
<td>51.25</td>
<td>46.63</td>
</tr>
<tr>
<td>Median time per clause</td>
<td>50.5</td>
<td>34</td>
<td>39.5</td>
<td>44</td>
<td>74</td>
<td>61</td>
<td>50.5</td>
<td>52</td>
<td>51.5</td>
<td>44.5</td>
</tr>
</tbody>
</table>

To analyse further, we then looked into behaviour log generated by the tool. We surprisingly learned that 93% of clauses having selected for content type and key concept before writing content obtained an average rate from two coaches as 4 and higher. Furthermore, learners with an average low rated result (id#3, id#5 and id#10) decided to fill writing content before providing content information. Eventually, they averagely spent about 52, 66, 64% of time respectively for writing and the rest of time for selecting content type and providing key concept. These patterns indicated that clear thinking about what to express apparently affects a quality of writing. In addition, a declaration of idea/thought can help to scope one thought and should be done before writing.

In the view of revision and improvement, we examined clauses that obtained a moderate rate and lower (under 3.0 rated) for analysis. The total clauses were 56 clauses from the tool and 73 clauses from not using the tool. All revisions made via the tool were found to be received all same rate or
improved rate. On the other hand, 26 from 219 of revisions from base 73 clauses from not using the tool (3 revisions limited) were worsen in rate while the rest was either the same or slightly improved. By asking coaches for a reason behind the worsen rate, they stated that the revision did not align with the comments and revised to another topic illogical and unrelated to surrounding context. Despite being guided by the same set of coaches, the decrease in rate implied that writing for novices without the strict environment to explicitly declare their thought can lead to confusion in thought. Moreover, unclear thought of a writer may bias and cloud a reception of incoming comments leading to interpret comments into another topic.

From interview after use, volunteered coaches mentioned that the tool impressively provided an insight detail of writing behaviour. The full log occasionally helped them to understand the learners thought and the cause of confusing expression in text. They gave an example that most times, when learners having trouble by spending a lot of time in selecting content type, the writing result would eventually become confusing and unclear. They also suggested that it could be more helpful if the tool can detect actions in depth such as there are whole-deletion for rewriting from a start or how long and how many times a learner does a consequence of ‘typing-pausing-deleting’. In the view of learners, we severally received a feedback that the tool helped them in collecting thought to prevent swaying in thinking. Seven learners mentioned that using the tool was a good practicing for becoming aware of thinking and collecting their thought since asking to provide content details prevented them to carelessly write without thinking. Six of learners also mentioned that they learned about importance of thinking/planning about goal.

4. Conclusion

This paper proposes a tool to assist in writing process for academic publication. The tool is designed as a strict environment for learners to become aware of their thought in writing process. To improve metacognitive skill in writing, explicit declaration of thought is asked to accompany each written clause. The tool promotes a collaborative interaction between learners and coaches to improve writing piece together. In usage, a learner as a main author provides a thought of what to express while a coach as a supervisor can instruct the flow of content and comment on content by considering thought and output. Moreover, the tool can generate a log including track change and behaviour in writing process. The behaviour can be used to hint about writer’s state of mind in writing. From an experiment, results implied that an environment of the tool helped to clear writers’ thought and improve their writing in terms of what to convey and logical connection. Additionally, behaviour log can help to track changes made for revision and study for writing actions to understand the root of misconception from a writer.

To improve the tool, we plan to include more detection on users’ actions on the tool such as counting on amount of deleting action and pause time while typing. These details can help a supervisor to learn more on a thinking process of learners; therefore, they can be suggested accordingly for sustaining improvement. Moreover, we plan to generate a flow of thought from the entire section to represent a connecting of thoughts from a writer. These connected thoughts can help to define another level of writing quality and can also be used to demonstrate an exemplar of good writing in logical connection among section from existing publication.

References

An Artificial Intelligence Approach to Identifying Skill Relationship

Tak-Lam WONG*, Yuen Tak YU, Chung Keung POON, Haoran XIE, Fu Lee WANG & Chung Man TANG

*Department of Computing Studies & Information Systems, Douglas College, Canada
bDepartment of Computer Science, City University of Hong Kong, Hong Kong
cSchool of Computing and Information Sciences, Caritas Institute of Higher Education, Hong Kong
dDepartment of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong
*tlwong@ieee.org

Abstract: Designing a good curriculum or an appropriate learning path for learners is challenging because it requires a very good and clear understanding of the subjects concerned as well as many other factors. One common objective of educational data mining and learning analytics is to assist learners to enhance their learning via the discovery of interesting and useful patterns from learning data. We have recently developed a technique called skill2vec, which utilizes an artificial neural network to automatically identify the relationship between skills from learning data. The outcome of skill2vec can help instructors, course planners and learners to have a more objective and data-informed decision making. Skill2vec transforms a skill to a vector in a new vector space by considering the contextual skills. Such a transformation, called embedding, allows the discovery of relevant skills that may be implicit. We conducted experiments on two real-world datasets collected from an online intelligent tutoring system. The results show that the outcome of skill2vec is consistent and reliable.

Keywords: skill relationship, neural network, learning analytics, artificial intelligence

1. Introduction

Designing a suitable learning sequence or learning path is essential for learners to effectively and efficiently acquire multiple knowledge or skills (Chen, 2008). Taking learning mathematics as an example, it is natural for learners to learn in the order of addition, subtraction, multiplication and then division due to the relationship between these concepts and the skills involved. Recent studies showed that learning trajectories, that is, the paths or sequences of learning activities involved to achieve a learning goal, can have significant impact to mathematics learning (Clements & Sarama, 2004; Schoenfeld, 1992). A coherent cumulative learning sequence can help learners better conceptualize the knowledge and lead to higher order thinking (Meyers & Nulty, 2009). Curriculum designers, educators and researchers have been paying great effort in reviewing the knowledge sequence of different subjects, or the curriculum across subjects in a formal learning context (Datnow, Borman & Stringfield, 2000; See, Gorard & Siddiqui, 2017). For example, the “Core Knowledge Sequence” adopted in the United States “identifies and describes the specific core of shared knowledge that all children should learn in U.S. schools. The Sequence is intended to help children establish strong foundations of knowledge grade by grade from preschool through grade eight.” In essence, a curriculum defines the sequence of knowledge and skill learning.

In a formal learning context, teachers can schedule the sequence of knowledge or skills to be learned by students based on the curriculum designed by the authority (such as state government). Teachers may also be familiar with students’ background and capability to plan and refine the schedule of teaching based on the students’ ability to assist students to learn more efficiently and effectively. However, teaching experience is subjective and various factors, including the groups of students taught, can affect it. Therefore, teachers may have different perspectives and preferences regarding the learning sequence for different skills. Taking mathematics as an example, factors such as culture and gender differences may influence the learning sequence of different topics and, hence, the curriculum design...
(Burny, Valcke, Desoete & Van Luit, 2013; Hyde & Mertz, 2009). This can pose additional difficulties in planning the learning sequence. It may also pose difficulties in planning a cross-subject curriculum, which involves the challenge of understanding the details of different subjects.

Thanks to the rapid development in information technology, a massive amount of data can be readily collected for analysis. For example, many universities and schools use course management systems such as Blackboard, moodle, and schoology for sharing learning resources, conducting online activities, attempting online tests, etc. These course management systems are able to keep track of the learners’ online activities for analysis and better understanding their learning (Romero, Ventura & Garcia, 2008). Similarly, many learning data such as the duration that the learner spends on viewing a video can be collected from massive open online courses (MOOCs). These learning data are very useful and valuable for course planners, lecturers, instruction designers, and learners themselves to enhance learning. Recently, Learning Analytics (LA) have been intensively researched to analyze the data about the learners and their contexts. The well-adopted definition of learning analytics is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.” (LAK, 2011) LA can provide instructors and learners with analytical information of the latter’s learning progress and hence lead to data-driven and well-informed decisions to optimize the learning.

This paper follows up the preliminary study reported in (Wong, Xie, Wang, Poon & Zou, 2017) and investigates the application of learning analytics technique to identify the relationship between the knowledge or skills based on the learning data available. The relationship between different skills can be readily visualized and easily interpreted by human. We carefully verified our approach by conducting experiments using two real-world datasets of learning algebra in an online intelligent tutoring system (ITS). The rationale of our approach is to automatically analyze the sequences of tasks that students undertake. Tackling each task requires the application of a skill. Therefore, a task sequence can also be interpreted as an evidence of the application of a sequence of skills. We aim at identifying the relationship between different skills from these sequences. To achieve this aim, we have developed an approach called skill2vec based on artificial neural network. In essence, our approach utilizes a two-layer neural network to transform an input skill to its contextual skills. In our approach, contextual skills refer to the neighbours of a skill in the sequence. The parameters of the neural network are “trained” automatically based on the available data. After training, our approach can essentially find a mapping from a skill to a vector in a new vector space that preserves the structure between different skills. This mapping, called an “embedding,” allows the identification of the relationship between skills by projecting the skills onto the new vector space. The results of an embedding can also be visualized by further projecting the vectors in the new space onto a two-dimensional space. We conducted experiments on datasets concerning student performance on mathematical problems from an intelligent tutoring system (Stamper, Niculescu-Mizil, Ritter, Gordon & Koedinger, 2010). For example, we discovered from the data that learning “Consolidate variables” then “Combine like-terms” in solving equation was similar to learning “Correctly placing points” then “Setting the y-intercept” in drawing an equation in a coordinate system.

2. Methodology

2.1 Problem Definition

The aim of this work is to identify the relation between skills by projecting a skill onto a lower dimensional space through embedding. Suppose there is a sequence of skills denoted by \((S_1, S_2, \ldots, S_T)\), where \(S_t\) is the skill required to tackling a particular task in the \(t\)-th step and \(T\) is the length of the sequence. For example, Figure 1 shows a sample question from an online assessment system. This question lists the three steps involved in answering the question, as follows: 1. [Triangle perimeter] finding the perimeter of the triangle in terms of \(k\); 2. [Consolidate variables] identifying terms with same variables; 3. [Combine like-terms] adding up the terms with \(k\). Each step is associated with a skill (or knowledge component), which refers to the information that can be used to accomplish the task. In this example, the length of the skill sequence is three and it consists of three different skills. Suppose there are altogether \(N\) different possible skills. Then each skill \(S_t\) can be
represented by a one-hot vector of length $N$, which contains only one element with value 1 and other elements with value 0. The vector is too sparse and does not have sufficient information from computing similarity between skills in this original vector space. To address this difficulty, the objective of our approach is to find a transformation (that is, an embedding) of this vector from the original $N$-dimensional space to a new $H$-dimensional space where $H << N$. The new space should be more expressive and be able to capture the context of the skills such that the relevant skills are close to each other in the embedding.

![Figure 1](image1.png)

**Figure 1.** A sample question in the dataset extracted from www.assistments.org.

### 2.2 Skill2vec

Our technique, *skill2vec*, borrows the idea of the SkipGram model, which achieves promising results in Natural Language Processing (NLP) problems (Mikolov, Sutskever, Chen, Corrado & Dean, 2013). One sample application is analogy reasoning, which can accurately infer professor given university, doctor given hospital, etc. Figure 2 depicts an overview of our *skill2vec* model. Essentially, it is a fully connected neural network model consisting of $H$ neurons in the hidden layer. The output layer contains a number of neurons depending on the predefined window size $W$, which is equal to 2 in our example.

![Figure 2](image2.png)

**Figure 2.** The *skill2vec* model.

The rationale of *skill2vec* is that for each skill $S_t$ in the sequence, the contextual skills are defined as skills for solving the neighboring tasks within a window of size $W$. The architecture of the neural network outputs the values corresponding to the contextual skills by inputting the skill $S_t$. In this way, the $N$-dimensional one-hot vector, $S_t$, which represents the skills needed to tackle step $t$, serves as the input of the *skill2vec*. This input vector will be fed to the hidden layer. The input to the $j$-th neuron of the hidden layer is defined as a standard weighted sum in the following manner:

$$ h_j = \sum_{i=1}^{N} x_i w_{i,j} $$
where \( x_i \) and \( w_{i,j}^1 \) refer to the input value of the \( i \)-th element of \( S_i \) and the weight connecting from the \( i \)-th element of \( S_i \) to the \( j \)-th neuron of the hidden layer. The superscript of \( w \) denotes that it is the weight to the first (hidden) layer of the neural network. The output layer consists of neurons that, together, show how likely it is for a particular skill to be among the contextual skills of \( S_i \). The output value of each output neuron is computed using the softmax function as follows:

\[
u_k = \sum_{j=1}^{H} h_j w_{j,k}^2, \quad y_k = \frac{\exp(u_k)}{\sum_{k'=1}^{S} \exp(u_{k'})} \]

where \( y_k \) is the output value of the \( k \)-th neuron of the output layer and \( w_{j,k}^2 \) is the weight connecting from the \( j \)-th neuron of the hidden layer to the \( k \)-th neuron of the output layer. Essentially, the values of the output neurons represent the probability that a particular skill is the contextual skill of the input skill. On the other hand, the hidden layer of \textit{skill2vec} serves as an embedding, which projects an input skill onto a new vector space in \( \mathbb{R}^H \). Technically, \( h \in \mathbb{R}^H \) is the product of the input vector \( x \in \mathbb{R}^N \) and the weight \( w^1 \in \mathbb{R}^{N \times H} \). As a result, two different skills can map to two similar vectors in \( \mathbb{R}^H \) if they have similar contextual skills.

![Figure 3](image)

**Figure 3.** The generation of training examples from a sequence of skills. The shaded and unshaded skills refer to the input and output of our \textit{skill2vec} model during training.

The weight of the parameters of \textit{skill2vec} can be trained or tuned automatically from the available sequences collected. For each available sequence of skills, training examples are created to train \textit{skill2vec}. Figure 3 shows an example of the generation of training examples from a sequence containing eight skills. Let the window size be 2. For each of the skills in the sequence, a training example will be created along with the neighbouring skills within the window size as illustrated in the Figure. Auxiliary skills will be created for the skills at the beginning or at the end to complete a training example. With respect to the Figure, the shaded skill and unshaded skills in each training example become the input and the output of \textit{skill2vec} during training respectively. Note that each skill is an \( N \)-dimensional vector. Given an input skill, the difference (or error) between the predicted output by \textit{skill2vec} model and the actual value of the output will be measured. The weights (i.e., \( w^1 \) and \( w^2 \)) of \textit{skill2vec} are then updated using standard backward propagation and delta rule. This process will be repeated by using the next available training examples and until the total error among all training examples is minimized or below a certain predefined value.

### 2.3 Experiments

We conducted experiments to demonstrate the effectiveness of our proposed method. First, we collected a publicly available dataset that included the interaction of 1,840 and 575 students in an ITS when they learned algebra in the academic years 2005–2006 and 2006–2007, respectively (Stamper, Niculescu-Mizil, Ritter, Gordon & Koedinger, 2010). We call these two datasets \textit{Algebra-2005-2006} and \textit{Algebra-2006-2007}, respectively. In the ITS, a learner performed a sequence of tasks to answer a given question similar to the sample question in Figure 1. The total numbers of steps for all students...
involved in Algebra-2005-2006 and Algebra-2006-2007 are 813,661 and 2,289,726, respectively.

We pre-processed the datasets by forming sequences of skills; each sequence being generated from a single question. Training examples were then generated from these sequences as described in Section 2.2. Training examples from these two datasets were used for training two skill2vec models separately. These two models, one from Algebra-2005-2006 and one from Algebra-2006-2007, served to triangulate the results of our model. The vector corresponding to each skill transformed to the new vector space by the embedding was then computed as described in Section 2.2. Next, the skill vectors in the new vector space were projected onto a two-dimensional space for visualization and analysis.

3. Results

Figures 4 and 5 show the visualization of parts of the results of our experiments on the datasets Algebra-2005-2006 and Algebra-2006-2007 respectively. The actual output of skill2vec contains hundreds of vectors related to all skills. For simplicity of exposition here, we only select and display 8 different related skills in Figures 4 and 5 to illustrate the usefulness and effectiveness of skill2vec. As described in Section 2.3, the vectors corresponding to the skills were projected on a two-dimensional space for better visualization. In each of the Figures, a point refers to a particular vector and its label shows the skill label associated with the vector. The arrows were added manually to connect the relevant skills in the results. For example, the skills “S1: Consolidate variables” and “S2: Combine like-terms” refer to two different, but related skills in algebra.

![Figure 4](image1.png) **Figure 4.** The results of applying skill2vec to the dataset Algebra-2005-2006.

![Figure 5](image2.png) **Figure 5.** The results of applying skill2vec to the dataset Algebra-2006-2007.

We found that the same skills are located in similar positions in Figures 4 and 5. The differences (or directions) between two different related skills are also similar. For example, “S2: Combine like-terms” is located at the southeast of “S1: Consolidate variables”, and “S3: Identify units” is also located at the southeast of “S4: Changing axis bounds” in both figures. However, “S5: Entering the slope” and “S6: Variable in denominator” are in the middle of Figure 4, but they are at the top of Figure 5. Recall that the two skill2vec models were trained using two different datasets Algebra-2005-2006 and Algebra-2006-2007 and the training involved randomness in neural network initialization. Hence, the two embeddings (and the new vector spaces) would be expected to be different unless they have some intrinsic relations. Nevertheless, the vectors in both embeddings and the difference between related vectors are largely consistent.

4. Conclusion

We have developed an approach called skill2vec for automatically identifying skill relationship from learning data. Skill2vec utilizes a two-layer neural network architecture whose input and output are a skill in a sequence and its neighbouring skills respectively. Skill2vec can find an embedding, which is a mapping transforming a skill from its original vector space to a new vector space. The new vector space is more expressive and related skills can be identified by considering the vectors in this new
space. We conducted experiments on two different real-world datasets collected from an online intelligent tutoring system to demonstrate the effectiveness and reliability of our approach.

Acknowledgements

The work described in this paper is fully supported by the grant from Research Grants Council of the HKSAR (Ref.: UGC/FDS11/E02/15).

References


Development of a System to Construct Explanation for Physics Phenomenon through Thought Experiment

Miki MATSUMURO*, Kazuhisa MIWA* & Yuiko TOMBEb
*Graduate School of Informatics, Nagoya University, Japan
bGraduate School of Information Science, Nagoya University, Japan
*muro@cog.human.nagoya-u.ac.jp

Abstract: This study focused on thought experiments (TEs) as a strategy to modify students’ incorrect internal mental models and aimed to develop a system to externalize TEs in students’ mind. A TE is an experiment conducted in the mind. Our system, called Thought Experiment Externalizer (TE-ext), provided some situations where TEs were conducted. The students externalized the results of their TEs on a PC display using the TE-ext. We tried to modify the student’ incorrect internal model and their incorrect explanation of surprising data using the system. The students were able to improve their explanation, which means that they were able to modify their internal models.

Keywords: Thought experiment, explanation, science education system, physics

1. Introduction

When we try to understand the world or solve problems, we construct an internal representation of an abstract world or the problem situation, called an internal model (Anzai & Yokoyama, 1984). Students often construct incorrect internal models based on their daily experience (Anzai & Yokoyama, 1984; Brown, 1993; Vosniadou & Brewer, 1992). We focused on thought experiments (TEs) as a strategy to modify students’ incorrect internal models.

A TE is an experiment conducted in one’s mind. The first step of a TE is to imagine a situation where an experiment is conducted (Kösem & Özdemir, 2014; Reiner, 2006). In the context of education (e.g., Clement, 1993; Smith, 2007), teachers give the most effective TE scene. TE users then carry out operations, such as pushing, pulling, or dropping an object. Finally, they see the results of the operation in their minds. The result is derived from their knowledge or memories, or from mental simulations (Clement, 2009; Hegarty, 2004). Although TEs do not provide new knowledge or information, it invokes knowledge unused in the original problem. In this paper, we introduce our system called a Thought Experiment Externalizer (TE-ext), which helps the students modify their internal models and construct a correct explanation for a physical phenomenon.

2. Thought Experiment Externalizer

2.1 Material

We used the yoyo problem from Anzai and Yokoyama (1984). The problem was to predict the movement of the yoyo in Figure 1(a). The correct answer was that the yoyo would rotate clockwise and roll to the right. The direction of rotational momentum caused by the tension force and the center of rotation at the yoyo’s point of contact with the floor determined the movement. However, according to the students’ naïve model, the yoyo’s axis was the center of rotation. They answered that the yoyo would rotate counterclockwise and roll to the left.
2.2 System’s Function

The TE-ext is a tool externalizing students’ TEs on a PC display. The display is divided horizontally; a specific TE scene is given above and the original yoyo problem scene is below (Figure 2(a)). The students pull the string as shown in Figure 2(b) and imagine how the yoyo will move (the yoyo on the display does not move at this point).

The students can move the yoyo in the display by selecting directions and pulling the string again (Figure 2(b), 2(c)). The yoyo in TE-ext moves in accordance with the selected buttons, even if such movement violates physical laws. The role of TE-ext is only to externalize TEs; in a mental simulation, any prediction is possible. Then the students consider why the yoyo moves as they imagined, drawing arrows on the display (Figure 2(d)).

2.3 Scenario

A TE scene is given according to a scenario consisting of three TE scenes in Figure 1. In the axis scene (Figure 1(b)), we put a stick through the center of the yoyo and set it between two sticks fixed on a table. This was a familiar scene to the students, reminding them of such as a toilet paper holder. They would easily imagine the yoyo rotating counterclockwise from their daily experiences. Then we fixed the yoyo to the sticks in the fixed scene (Figure 1(c)). The yoyo would not move as the string was pulled.

The broken scene (Figure 1(d)) was the successor to the fixed scene. The instructions for this scene were as follows: “You have kept pulling the string of the fixed yoyo. As a result, the sticks almost break. If you pull the string again, they will break completely.” This scene was identical to that of an object standing on a table being pulled to the right. The yoyo and sticks would fall down to the right, rotate clockwise and move to right.

3. Test of the effect of the TE-ext

3.1 Participants and Procedure

Twenty-three undergraduate students participated. First, we demonstrated the movement of the yoyo in the original problem. Following the demonstration, the students gave an explanation as to why the yoyo moved as shown in the demonstration as the pre-test.
After the pre-test, the students considered why the yoyo in the original problem rotated clockwise and moved to the right using the TE-ext for 30 minutes. After the thinking time, the students saw the demonstration and gave an explanation again as the post-test.

3.2 Results

Four students were excluded from the analysis, because they did unrelated things. The students’ explanations were scored from 0 to 3 point. If the students continued to work with the incorrect internal model, their explanations were scored 0 point. We gave 1 point, if they related the direction of the pulling the string and the movement of the yoyo. For example, “the experimenter pulled the string to the right, so the yoyo moved to the right and rotated clockwise.” A score of 2 points was given to the explanations that described the pulling tension affecting the whole of the yoyo. Explanations implying that the student understood that the yoyo rotated clockwise because its point of contact with the table did not move (i.e., the center of the rotation) were given the score of 3 points. The score in the post-test was higher than that in the pre-test significantly (pre $M = 0.684$, $SD = 1.003$ vs. post $M = 1.368$, $SD = 1.012$; $t(18) = 2.233$, $p = .038$).

4. Discussion

The explanations of the students were improved using the TE-ext. They were able to modify their internal model. They learned the forces exerted on the yoyo from each scene and applied this knowledge to the original problem. Some might think that the students would always modify their internal models if they were given surprising results. However, previous studies have shown that even if they receive a surprising result, students resist modifying their models (Chinn & Brewer, 1993). Almost half of the students (57.89%) tried to make explanations without modifying their internal model (score 0) in the pre test in this study.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP 15H02927.

References


Towards A Virtual Peer that Writes Stories with Children

Hans Gabriel CHUA, Geraldine Elaine CU, Chester Paul IBARRIENTOS, Moira Denise PAGUILIGAN & Ethel Chua Joy ONG

Center for Language Technologies, De La Salle University - Manila, Philippines
*ethel.ong@delasalle.ph

Abstract: Children encounter more difficulty in writing than in sharing their stories verbally. This difficulty may be due to the need for them to organize their thoughts and express them in the written language. A software agent can be designed to mimic the support given by a human teacher or peer, such as asking questions, giving directives, and recommending possible story text. We describe how these tasks can be achieved by our software agent, Alice, using a combination of computational text understanding and generation techniques. Children found Alice's prompts to be useful in situations when they were having a hard time writing their story. Those children who opted to include Alice's suggested story text did so because they found the peer to state certain facts that can add new details to their stories.

Keywords: Virtual Peer, Story Writing, Text Understanding, Text Generation

1. Introduction

A common practice in the classroom setting is to ask children to write stories, usually on a variety of topics. This presents them with an opportunity not only to build their vocabulary and grammar, but also to express their ideas while constructing and extending their knowledge on a particular area. Their decision-making and creativity skills are enhanced as well when they think about what can happen next in the story. However, children usually encounter more difficulty in meeting the demands of writing stories than its verbal counterpart. The problem lies in having to organize their thoughts, the need to use appropriate language to communicate their ideas through the written text, and their lack of knowledge about writing and the writing process (Englert, 1988). Verbal storytelling also affords children with an opportunity to engage in a fun activity with peers and caretakers, whereas a writing task usually involves individual work.

Research in virtual agents have designed them to communicate with their human users through different modalities and interfaces, including written text and spoken dialogue, facial expressions, and body gestures. Interaction is promoted through creative collaborations with the user through social engagements and communication, which are vital in literacy learning (Cassell, 2004), language development (Cassell et al., 2005), reasoning (Blair, 2006), and critical thinking. Advances in natural language processing research have also extended the capabilities of virtual agents to enable them to mimic natural conversations with human users (Ong, 2016). In this paper, we describe our virtual peer, named Alice, who collaborates with a child in the story writing task. Alice uses natural language processing techniques to understand a partial story text and to generate responses.

2. Designing Alice

Alice supports children in writing stories by serving two roles – facilitator and collaborator, using the story writing space shown in Figure 1. Instead of a turn-taking approach, the child user is given control to decide when to ask Alice for help ("Ask Me" button), and the type of help needed from the virtual peer (Figure 2). However, Alice imposes certain restrictions on the story being written. The story plot must follow the three-act-structure, as depicted in Figure 3.
The three-act structure divides a story into three main parts - the beginning, the middle and the end. The beginning part introduces the characters and the location. The conflict is also revealed in this part. The conflict is the problem faced by the characters that creates tension and makes the plot move forward (Cox, 2016). In children's stories, a conflict can be depicted with a story text that introduces a negative emotion or tension. Alice detects the presence of a conflict by looking for a concept that produced the least negative polarity, with the use of SenticNet (Cambria et al., 2014).

The middle part focuses on narrating a series of events involving the protagonist that may lead him/her to the resolution of the conflict. At least two events must be detected. Events are identified by the use of verbs that denote actions, such as eat, play and read.

The end part is the climax that narrates how the character resolves the conflict. At least one of the characters that caused the conflict must experience the resolution. Resolutions are determined by looking at two factors. First, a resolution must be related to the conflict previously introduced in the beginning part of a story. Second, the resolution has a positive polarity. The commonsense ontology is used to determine if the resolution concept is related to the conflict concept. Related means the path from the conflict concept to the resolution concept contains no more than three (3) edges. Finally, the resolution must also be experienced by the character who was involved in the earlier conflict.

Alice can generate two types of responses: story text or dialogue. Dialogues are prompts in interrogative or imperative form to encourage the user to provide more details about a character, an object, a location, or an event. In story text generation, Alice needs to provide additional details to the nouns or events that it has previously detected from the user's story.

3. Results and Discussion

Alice was validated through two iterations of end-user testing comprising of 24 students as participants, who are between the age of seven to nine years old. Each participant used Alice to write his/her story. The participant’s stories and Alice’s responses were internally logged for further analysis. A debriefing interview was conducted with each of the participant to solicit feedback regarding their experience.

52.9% of the participants claimed that they were able to finish writing their stories. However, an analysis of the internal logs showed that none of them actually did. This is because Alice failed to detect some story elements, such as location, conflict, character and resolution. Part of the problem lies in Alice's reliance on Stanford CoreNLP's named entity recognizer, which did not treat a common noun as a possible story character or a location. The frequency with which Alice failed to detect the story elements is summarized in Table 1.

The presence of a sentence with any concept that has a negative polarity is used to identify a conflict. However, most participants in the first iteration describe their experiences without using any words bearing a negative sentiment. During the second iteration of testing, most children wrote a story
whose theme centers on what the main character "hates". This conflict can then be easily resolved by writing a sentence containing the word "likes". Thus, 28.6% of the participants wrote stories that Alice deemed as complete. On the other hand, 42.9% were able to reach the end part of their stories, but were not able to give a resolution that is “acceptable” to Alice's criteria.

Table 1: Frequency of missing story elements

<table>
<thead>
<tr>
<th>Missing Story Element</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>15</td>
</tr>
<tr>
<td>Conflict</td>
<td>11</td>
</tr>
<tr>
<td>Character</td>
<td>4</td>
</tr>
<tr>
<td>Resolution</td>
<td>17</td>
</tr>
</tbody>
</table>

In the usability aspect, 91.7% of the overall participants asked Alice for help at least once during their writing session. Their reasons for doing so include the following: (i) they were having a hard time in writing their story; (ii) they want to know more options on how their story may happen; and (iii) they find Alice's prompts and suggested story segments amusing. Participants who already have a story in mind claimed that they did not need the help of Alice, so they did not ask for help.

4. Further Work

In this research, we applied text understanding and text generation techniques to enable Alice, a virtual agent, to facilitate and even co-author with a child while the latter is engaged in story writing. We showed how Alice afforded opportunities for each child to write at his/her own space, using a theme that is of interest to him/her, and to work with a peer as deemed necessary by the child.

Story writing touches on many aspects, from general themes that may apply to all children, to specific stories that narrate a child's personal experiences. Alice should be able to possess a vast domain of knowledge to cater to these individual differences. Alice can be designed to utilize story starters or prompts. These ready-made scripts on common themes and familiar life events can give children an initial idea that they can build on in their stories.

The internal logs also showed that the stories written by the children contained numerous spelling and grammar errors. Alice currently does not address linguistic concerns as part of its scope. Future research should extend Alice to assume the role of a language teacher in order to correct these errors. Furthermore, Alice can use this opportunity to enrich the vocabulary of the individual writer by suggesting various synonyms, using advanced words when generating a suggested story text, and demonstrating the differences of word usage.

References


Stanford CoreNLP. Available at https://stanfordnlp.github.io/CoreNLP/
Prospects in Modeling Reader’s Affect based on EEG Signals

Kristine KALAW*, Ethel ONG & Judith AZCARRAGA
College of Computer Studies, De La Salle University, Manila, Philippines
*kristine_kalaw@dlsu.edu.ph

Abstract: Readers experience various emotions while reading, which may affect their overall enjoyment and comprehension of the material. The current work presents a study on brainwaves or EEG signals and their association to emotions while a person is reading literary fiction. EEG data from 32 participants, aged 18 years old and above, were collected with the use of an EEG headset. We describe our methodology for data acquisition and processing, feature extraction and dataset building, as well as the classification experiments done.

Keywords: EEG, affect recognition, machine learning

1. Introduction and Challenges in Building the Reader Affect Model

People experience a variety of emotions while being engaged in certain activities. As shown in the studies of Azcarraga & Suarez (2012) and Yazdani et al. (2012), emotions evoked while being engaged in an activity could be detected via a person’s brainwaves or electroencephalogram (EEG) data. The act of reading literary fiction is a profoundly emotional experience. Miall & Kuiken (1994) and Cupchik et al. (1998) are empirical works in the culture, media, and arts that have established the relation between reading and emotional response. The current work presents an EEG-based detection and recognition of emotions while a person is reading literary fiction.

Collecting the data, obtaining the ground truth, and defining the appropriate emotion model are some of the challenges in building any affective model (Picard, 2003). For this study, we used an Emotiv Insight EEG headset (https://www.emotiv.com/insight/), which can capture data from the AF3, AF4, T7, T8, and Pz channels, with the convenience of a dry sensor. In obtaining the ground truth, a self-reporting scheme was integrated with the developed data collector tool. For the emotion model, two models were considered. The Hourglass of Emotion (HoE) model by Cambria et al. (2012) describes what is the person feeling. They described their model as one that is able to represent affective states both through its labels and its 4 independent but co-occurring affective dimension—namely, aptitude, attention, pleasantness, and sensitivity. In this way, their model can potentially describe a full range of emotional experiences. Whereas, the Emotions of Literary Response (ELR) model by Miall & Kuiken (2002), describes if the emotions were caused by overall enjoyment in reading the text (evaluative feelings), by the events or characters in the fictional world (narrative feelings), or by the formal components of the text (aesthetic feelings).

2. Methodology

2.1 Data Acquisition and Processing

EEG signals from 32 participants of various demographics were collected while they were reading The Veldt by Ray Bradbury. Prior to the experiment proper, the participants were asked to close their eyes and relax for 2 minutes. This recording serves as the baseline. Following the experimental set-up of Miall & Kuiken (1994) for the presentation of the stimuli, the story was manually divided into 72 segments using phrase and sentence divisions while still retaining meaning and coherence by itself. The story segments were presented one by one via the data collector tool adapted from Azcarraga & Suarez (2012). After reading the segment, the participant would annotate what they felt and what caused it. The
read-annotate process is repeated until the last segment is reached. Note that the EEG recording while the participants are reading is called the event-related potentials (ERP).

Data preprocessing includes synchronizing the ERP recording with the emotion annotations, dividing the data into the corresponding story segments, and then further subdividing each segment into 2-second windows with 1-second overlap. The total number of these windows represents the number of instances for a particular participant.

2.2 Feature Extraction and Dataset Building

EEG frequency bands were extracted via a series of band pass filters. For each band, the signals were transformed to the frequency domain using a fast Fourier transform algorithm, and then the minimum, maximum, and mean values for each feature type were computed. The feature types are magnitude, which is the absolute value of the signal; power spectral density (PSD), which is the square of the signal in the frequency domain; and spectral power of asymmetric electrode pairs via differential asymmetry (DASM), or power subtraction, and rational asymmetry (RASM), or power division. This makes a total of 252 extracted features for each instance.

Each instance is labeled according to the self-reported emotion annotations the participants made. With regard to the HoE, the participants were asked to rate each of the 4 dimensions from -3 to +3. If the value is negative, the assigned label is low. If the value is positive, then the assigned label is high. With regard to the ELR, whichever the participant chose is the assigned label. It is possible for the instance to have multiple assigned labels for ELR.

The current work is concerned with the brainwaves of participants during reading time. These brainwaves (ERP) are isolated by employing baseline correction, which is simply subtracting the ith feature value of the baseline from all the ERP instances (Woodman, 2010). After performing baseline correction and building the datasets according to the profile groups, the datasets are standardized by computing for the z-score values. Extreme z-score values were clipped to at least -3 and at most +3.

3. Preliminary Results

The experiments conducted are binary classifications of the HoE and ELR models on different datasets (Female, Male, Sex-merged) and classification methods (Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perceptron (MLP)). Across all classification experiments, the classifiers were trained with leave-one-participant-out cross-validation. Refer to Table 1 and Table 2 for the listing of averaged f-measure values of the classification experiments.

Table 1: F-measure values for the HoE class labels.

<table>
<thead>
<tr>
<th>DT (F)</th>
<th>DT (M)</th>
<th>DT</th>
<th>SVM (Base)</th>
<th>MLP (Base)</th>
<th>SVM (PCA)</th>
<th>MLP (PCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.28</td>
<td>46.12</td>
<td>49.53</td>
<td>46.52</td>
<td>52.04</td>
<td>47.51</td>
<td>48.86</td>
</tr>
</tbody>
</table>

The first experiment was to set the baseline performance using Decision Trees and to see whether there is an improvement in performance among the specific datasets and the general dataset. It is observed that on an average basis, there is no significant improvement in the performance between the general profile dataset and the specific datasets (refer to columns 1 to 3).

The next experiment was to attempt to improve the classification performance with SVM and MLP classifiers. As discussed in the previous section, there is no significant improvement in the performance between the general profile dataset and the specific datasets; therefore, the general dataset is the one used in this experiment. It is observed that MLP yields a slightly better performance than DT (refer to columns 3 to 5).

The last experiment was to see whether reducing the number of features via PCA could yield a result that is higher than or at least at par with the base feature set. It is observed that the performance of the classifiers with PCA feature sets yield subpar results to that of its counterpart with the base set of 252 EEG features. The average difference in the f-measure value of the Base classifiers and PCA classifiers for both SVM and MLP is ±2 (refer to columns 4 to 7). Note that the processing of the SVM and MLP classifiers takes much of the computer’s resources. Thus, if it is acceptable to have a ±2 margin of error, then using the PCA feature set would suffice as compensation for faster processing time.
Table 2: F-measure values for the ELR class labels.

<table>
<thead>
<tr>
<th>DT (F)</th>
<th>DT (M)</th>
<th>DT</th>
<th>SVM (Base)</th>
<th>MLP (Base)</th>
<th>SVM (PCA)</th>
<th>MLP (PCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.38</td>
<td>45.22</td>
<td>45.32</td>
<td>44.25</td>
<td>52.01</td>
<td>43.15</td>
<td>49.60</td>
</tr>
</tbody>
</table>

4. Future Work

We describe a methodology for building a reader affect model based on EEG signals collected from 32 participants. Inferences from our preliminary results are consistent with the emotion study of Kret and De Gelder (2012); however, further research and analysis is recommended.

Further work will consider strategies for improving the classification performance results, exploring the use of deep learning, and visualizing and showing the trajectory to discover patterns in reading behavior and preferences. Other experiments can explore intra-subject classification, wherein the same methodology may be repeated except that the stories that a single participant reads subscribe to one of the 6 core emotional arcs (Reagan, Mitchell, Kiley, Danforth, & Dodds, 2016). The current work makes use of the first impressions of the participants towards the story. Following what Tompkins (1980) said that reading is an experience that is never the same from one reading to the next, this could be tested by having the participant repeat the data acquisition process for a number of times. In this way, the fourth domain in the ELR model, self-modifying feelings, which involves the restructuring of the reader's understanding of the textual narrative, could potentially be mapped. The trajectory in the change of emotions for the same stimuli could be observed.

For intelligent tutors and embodied conversational agents, the resulting models can be used as basis for the conversation topic with the reader, to address factors affecting one’s engagement with the reading task and comprehension of the reading materials.

Acknowledgements

We would like to acknowledge the support and funding provided by the DOST-ERDT program.

References


Proposal of an Intelligent Tutoring System for Procedural Learning with Context-aware Dialogue

José PALADINES\textsuperscript{ab*} & Jaime RAMÍREZ\textsuperscript{b}
\textsuperscript{a}Technical Sciences Faculty, Universidad Estatal del Sur de Manabí, Ecuador
\textsuperscript{b}Computer Science School, Universidad Politécnica de Madrid, Spain
*Jose.paladines@unesum.edu.ec

Abstract: This poster presents a proposal for an ITS for procedural training in a 2D/3D virtual environment in what each student will be trained through a dialogue in natural language that takes into account his/her specific features, his/her progress in the development of the task and the environment where the task is performed. To support the dialogue, we will use a dialogue manager built on some of the platforms for creating dialogue managers currently available.

Keywords: Intelligent tutoring system, Natural language processing, Context-aware dialogue, Graphic virtual environment.

1. Introduction

Intelligent Tutoring Systems (ITS) have been developed for very varied environments and different purposes. However, there are very few ITSs for procedural training in which the student can interact with the system through an avatar in a 2D/3D virtual environment, and can dialogue with the system by using natural language.

Therefore, we present a proposal for an ITS for procedural training in a 2D/3D virtual environment with the capacity for dialoguing based on the context, that takes into account his/her specific features, his/her progress in the development of the task and the environment where the task is performed. So, tutoring feedback will be the result of a context-aware dialogue.

The dialogue manager will be built on some of the well-known platforms that have reached a remarkable degree of maturity for their facilities to interpret and process natural language in different application domains.

2. Related Work

Next we will present some of the most remarkable ITSs with Natural Language Processing (NLP) for 2D/3D virtual environments oriented to teaching. Why2-Atlas (VanLehn et al., 2002) supports that students write long explanations of simple mechanical phenomena, and uses a symbolic analysis (CARMEL) to discover their misconceptions. Autotutor (Graesser & et al., 2005) maintains mixed initiative dialogue in interactions in which students can explain concepts, and then their explanations are compared with a set of expectations and misconceptions by using Latent Semantic Analysis. Beetle II (Dzikovska, Steinhauser, Farrow, Moore, & Campbell, 2014) implements an approach based on task-oriented dialogue systems, and uses an ontology to represent domain knowledge.

On the other hand, the only ITSs with NLP for procedural training are Paco and Jacob. Paco (Rickel, Lesh, Rich, Sidner, & Gertner, 2002) supports tutoring actions as part of a collaborative dialogue system (built on Collagen) that uses rules based on a task model, a student model and the interaction with the student. Jacob (Evers & Nijholt, 2000) teaches to solve the problem of the Towers of Hanoi in a 3D virtual environment, and the interaction with the user is given by performing actions concerning the task. However, both Paco and Jacob suffer from many limitations regarding natural language dialogue.
3. Platforms for Creating Dialogue Managers

There are several platforms for the management of dialogue in natural language, between the most well-known are Api.ai from Google, Wit.ai from Facebook Messenger, Bluemix from IBM and LUIS from Microsoft. In general, these tools perform NLP within a workspace where the intentions, entities and the dialog flow are defined.

- Intentions represent the purpose of a user's entry (request).
- Entities represent a term or object that is relevant to intentions and provides a context for intention. Most platforms have predefined intentions that give a greater robustness to the application.
- The dialog is a logical tree-like conversation flow that defines how the application responds when it recognizes defined intentions and entities. Dialogue can be dynamically adjusted to the situations of a certain scenario by means of the context variables.

4. Features of an ITS for Procedural Training in a 2D/3D Virtual Environment with NLP

The dialogue manager has to implement a context-aware dialogue. To achieve this, the context should contain information on the virtual environment (static and dynamic), the student’s knowledge and progress in the activity to be performed. With this information it will be possible to provide a personalized tutoring through a context-aware dialogue adapted to: what the student knows; his/her avatar location in the virtual environment at the time of the dialogue; and the phase of the procedure in which the student is. Some examples of personalized tutoring feedbacks would be the following ones:

- Answer questions about the location of an object in the virtual environment or how to reach it, even when this object is far away from the student’s avatar.
- Answer questions about the next action to be done.
- Recommend learning activities to bridge knowledge gaps.
- Provide proactively hints to guide the student with the execution of a task.
- Encourage an affective dialogue in the face of student inactivity or moments of uncertainty.

5. Proposed Approach

In Figure 1 we show the architecture composed by three main components, which would provide a personalized tutoring centered around the virtual environment where the training is performed.

![Figure 1. Architecture of ITS with NLP for procedural training](image-url)

The procedural training environment is the scenario that simulates the real world where the training activities are carried out. The user will be able to interact with the environment generating events such as performed actions, actions attempts, questions, etc., which will be provided to the ITS throughout an exercise.
The **Intelligent Tutoring System** will integrate the modules corresponding to a classic ITS such as the student model, tutoring model, expert model, communication model.

Additionally, a world model will be added to the ITS to represent the characteristics of the training environment, this model will represent the physical structure of the scenarios and their content so that the system can answer a question regarding the situation of an object or how to go from one place to another.

Within the student model, the user's request will be evaluated through rules, which will determine what the student knows and doesn’t know and begin to structure the context. To this end, we will adopt the student model proposed in (Clemente, Ramírez, & de Antonio, 2011).

The context will be defined in terms of an ontology and it will be populated with information coming from the student model (knowledge state, activity progress, student trace), the expert model (correct plan, next correct action) and the world model (virtual world structure, student avatar position, object descriptions). Then, before passing the context to the dialogue manager, the ontological representation of the context will be translated into another representation understandable by the dialogue manager.

The **dialog manager** will contain the definition of the dialog structure, intentions and entities specifically intended for the training environment to be able to control the branching mechanism of a conversation. This component will be responsible of the communication with the user taking into account the contextual information provided by the models that integrate the ITS. Therefore, the hint given to the student will be adjusted to his/her current knowledge. Additionally, dialogue manager will be able to update the context after inferring the student knowledge from his/her utterances. This updated context will be sent back to the ITS.

### 6. Future Work

This is an ongoing work of a PhD thesis. In addition to what we have presented, in the future we plan to refine the presented architecture by: deepening in the concept of "context"; and designing a dialogue structure that leverages the context information to provide natural language messages properly adapted to the student and his/her situation in the virtual environment.

### Acknowledgements

Paladines would like to acknowledge financial support from the Universidad Estatal del Sur de Manabí.

### References


Improvement of Situational Dialog Function and Development of Learning Materials for a Japanese Dictogloss Environment

Satoru KOGURE\textsuperscript{a}, Kaito OKUGAWA\textsuperscript{a}, Yasuhiro NOGUCHI\textsuperscript{b}, Tatsuhiro KONISHI\textsuperscript{a}, Makoto KONDO\textsuperscript{a} \& Yukihiro ITOH\textsuperscript{b}

\textsuperscript{a}Faculty of Informatics, Shizuoka University, Japan
\textsuperscript{b}Shizuoka University, Japan
\textsuperscript{*}kogure@inf.shizuoka.ac.jp

Abstract: Dictogloss is a learning activity in which learners can cooperatively learn four language skills. We have constructed a dictogloss environment that is able to self-study for learners who study second language. As a previous system, we focus two problems. The first problem is that dialog function to communicate reasons why focused word is wrong is inadequate. The second one is that there are not adequate learning materials for language forms that include communication for reasons. In this paper, we report the solutions for two problems.

Keywords: Dictogloss, Japanese education, corroborative learner agent, self-study

1. Introduction

Dictogloss is a learning method in which learners can cooperatively learn four language skills; listing, speaking, reading and writing (Wajnryb, 1990). In dictogloss, learners learn the target language by the three stages: (S1) dictation stage, (S2) reconstruction stage and (S3) analysis and judgement stage. In S1, the teacher reads short sentences including the focused language forms. In S2, learners take notes about heard sentences and reproduces sentences that have same meaning of original sentences by consulting with collaborative learners. Finally, In S3, the teacher performs answer matching and explains the error of reconstruction sentences. Many studies focus on the effect using dictogloss activity for learning each language skill (Jibir-Daura, R., 2013, Sari Dewi, R., 2014, Lindstromberg, S., et. al., 2016). There is also research on a dictogloss learning system. However, the activity of dictogloss is not suitable for self-study because collaborative learners and teachers are necessary. Therefore, we have constructed a dictogloss environment that is able to self-study for learners who study second language (Kondo, M., et. al., 2012, Tashiro, A., et. al., 2013, Kogure, S., et. al., 2015, Kogure, S., et. al., 2016). In the dictogloss environment, we realize a cooperative learner agent (CLA) and teacher agent (TcA) on the system. In the system, the learner inputs the reproduction sentences using the keyboard. The system does not cover the speaking skill. The learner compares two reproduced sentences written by CLA and themselves. If the learner determines the reproduced sentence is wrong, the learner clicks on the word that they click on the wrong word in the reproduced sentences written by CLA. Then, the system automatically generates a message to CLA for pointing out mistakes (Kondo, M., et. al., 2012, Tashiro, A., et. al., 2013). Learners can also give the CLA a reason why the word is wrong (Kogure, S., et. al., 2015).

As an examination of our dictogloss system, we focus two problems as follows:
(Problem 1) Dialog function to communicate reasons why focused word is wrong is inadequate
(Problem 2) There is not adequate learning material for language forms (LF) that include communication for reasons.

The purpose of this study is to solve these two problems. For archiving these solutions, we construct adequate dialog function to communicate reasons (Solution 1) and prepare some enough learning materials (Solution 2).
2. Improvement of Dialog Function for Reason

A teacher could only use causality order and causality paradox as conjunctional LF in our previous environment. We extended the conjunction LF so that the teacher can deal with five relationships: conjunction reason, conjunction time, conjunction condition, conjunction paradox and conjunction purpose. We decided these six conjunction relations regarding 3A Corporation Ed. (1998) and 3A Corporation Ed. (2000). We eliminated various knowledge on the reason function of the existing system from program and improved the system so that a teacher gives the system the XML format files that include the information for the reason function.

We take Japanese Yari-morai expression as situation reason. Yari-morai expression is a combination of Yari (“give” in English) and Morai (“receive” in English), which is not in English. Yari and Morai have the role of auxiliary verb. This auxiliary verb adds a weak gratitude exchanges to the main verb. Japanese Yari-morai expression has three types: “shite-ageru” (T1), “shite-kureta” (T2) and “shite-morau” (T3). Table 1 summarize types of Yari-morai expression. For example, a T2 sentence “Watashi ha anata ni suugaku wo oshiete kureta” (I taught mathematics to you) is ungrammatically in Japanese. It is very difficult for beginner to understand Japanese Yari-morai expression. English expression cannot distinguish T1 and T2. T1 is used when the actor is themselves or a person close to themselves. T2 is used when the target person is themselves or a person close to themselves. T1 and T2 are in the active and T3 is in the passive. When generating errors for the above three types, the CLA selects one of different type expression. In case using generation rule, the system generates either grammatical sentences or ungrammatical sentences. The teacher gives the system in advance the knowledge of whether it is grammatically correct or not (3A Corporation Ed. (1998) and 3A Corporation Ed. (2000)).

Table 1: Types of Yari-morai expression for sample situation.

<table>
<thead>
<tr>
<th>Type</th>
<th>Active or Passive</th>
<th>Subject in sentence</th>
<th>Object in sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: shite ageru</td>
<td>Active</td>
<td>Person for agent</td>
<td>Person for target</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NG: me or family</td>
</tr>
<tr>
<td>T2: shite-kureta</td>
<td>Active</td>
<td>Person for agent</td>
<td>Person for target</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NG: I</td>
</tr>
<tr>
<td>T3: shite-morau</td>
<td>Passive</td>
<td>Person for object</td>
<td>Person for agent</td>
</tr>
</tbody>
</table>

T1 or T2: An agent person give anything to a target person.
T3: A target person is given anything by an agent person.

3. Expansion of Learning Materials

We construct the system to be able to indicate the error patterns by external data file to realize generic learning material configuration. Then, we created 19 learning materials dealing with contextual reason and 3 materials dealing with situational reason. We prepared 3 or 4 sentences for each of 22 materials. We think that materials are enough to understand the basis of Japanese for beginner according to refer to 3A Corporation Ed. (1998) 3A Corporation Ed. (2000). We also prepared 2, 3 or 4 mistakes for each focused LF regarding Ichikawa, Y. Ed. (2010). One of the authors uttered and recorded the content speech of the lesson sentence. We prepared a situation diagram according to search suitable images from free web site. We assume 3 minutes to listen the lesson sentences, 6 minutes to talk to CLA and 6 minutes to the feedback phase. Therefore, we estimated the time it takes to learn one learning material to be approximately 15 minutes. We prepared about 330 minutes learning contents for all 22 materials. Learners can learn about 5 hours and a half in total with newly prepared materials. We guess that we could have prepared enough materials for beginner.

4. Experimental Evaluation

The purpose of this experimental evaluation is to evaluate the ease using the dialog function for the reason. Subjects are four Japanese undergraduate students. We let the subject learn the learning
materials on the two conjunction relationships and two Yari-morai expression in the system. We conducted a questionnaire that include the following evaluation items after the experiment.

(1) For conjunction relationship: (a) Usability of the interface for indicating contextual wrong and (b) Nature of dialog sentences with CLA
(2) For Yari-morai expression: (a) Usability of the button for indicating contextual mistake, (b) Usability of the button for indicating grammatical mistake, (c) Usability of the interface for indicating situational mistake, (d) Usability of the interface for indicating grammatical mistake and (e) Nature of dialog sentences with CLA

We got a rating of 4.00 points or more in all items ((a) 5.00 and (b) 4.50 in (1) and (a) 4.75, (b) 4.55, (c) 4.50, (d) 4.75 and (e) 4.75 in (2)). From this result, it is suggested that there is no problem in usability of proposed interface.

We also got some opinion in free description. There was an opinion that it is difficult to understand whether the CLA error is due to situational mistake or grammar mistake. We believe that prior lecture on the focus LF form is necessary before using our system because there are situations that Japanese students cannot judge quickly. Furthermore, we think that it is necessary to implement hint functions, etc. while using the system.

5. Conclusion

We had expanded the reason dialog function for conjunction LF and implemented the reason dialog function for situational reason for solving problem 1 and we had expanded the learning materials for solving problem 2. We prepared about 330 minutes learning contents for all 22 materials. We got a rating of 4.00 points or more in all evaluation items from Japanese students. In future works, we will evaluate the system by abroad student who studies Japanese.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 25730204 and 26350273.

References

Semantically Enhanced Gaze-aware Historical Cartoons to Encourage Historical Interpretation

Daiki Muroya*, Yuki Hayashi & Kazuhisa Seta
Graduate School of Humanities and Sustainable System Sciences, OPU, Japan
*muroya@ksm.kis.osakafu-u.ac.jp

Abstract: While it is desirable that learners tackle their learning activity by trying to read between lines and give an interpretation to lighten the background context of the knowledge, most of them merely focus on memorizing what is taught to improve their test scores. In this research, we aim to develop a learning support system that encourages learners to change their learning attitudes to active ones through performing historical thinking activities. Our approach is to focus attention on the potential of historical cartoons as learning materials. In this paper, we discuss the concept of an adaptive learning support method by utilizing semantically enhanced historical cartoons.

Keywords: Historical thinking, Historical cartoons, semantic representation, historical interpretation learning

1. Introduction

The importance of cultivating learning attitudes that enable learners to not only memorize fundamental knowledge but also give an interpretation to lighten the background context of the knowledge is widely shared. However, since teachers using traditional instructional methods are often obliged to unilaterally teach basic knowledge to learners, most learners fail to understand that rote learning (memorizing what is taught) is not an appropriate learning technique. Consequently, they tend to just acquire shallow knowledge without organization and feel bored with learning.

In history learning as well, the importance of cultivating ‘historical thinking skills’ that require the abilities to read between the lines of historical events with considering historical background is known (Linda et al, 2012; Kenneth et al, 2000): For example, during “The Sino-Japanese War”, the victory of Japan, which was still a little-known country in those days, resulted in the acceleration of the Chinese continent colonization by several western powers, ultimately leading to the rising of the imperialism era. Such kind of interpretation of historical events’ background would be a desirable learning attitude among learners. However, most of them rather tend to focus on memorizing only names of battles, places, characters, events, etc., as a result of the lack of ‘viewpoints’ to give their historical interpretations to each event (Spoehr & Spoehr, 1994).

The objective of this research is to develop a learning support system whereby learners change their learning attitudes to active ones through performing historical thinking activities with given historical viewpoints as necessary for the historical events. To approach the objective, we use historical cartoons, which usually act as visual metaphor to illustrate historical interpretation, as learning materials so that learners perform historical thinking activities. In this paper, we propose an adaptive learning support method using historical cartoons with systematized semantic representation.

2. Design Philosophy of Historical Interpretation Learning

Illustration-based historical cartoons make it easier even for novice learners to intuitively understand the overviews of historical events in comparison with text-based historical literatures such as ancient documents and paleography. In the educational context, utilizing historical cartoons as learning materials has several pedagogical values. Cartoons, for instance, provide the learners with opportunities to gain insights into what people were thinking at that time, since they are often summarized and
encapsulate the social problems behind the historical events effectively based on the painters’ keen eyes (Stradling, 2001).

In order to develop a learning support system which encourages learners’ historical interpretation activities meaningfully, we employ a learning methodology proposed by Kuroda (Kuroda, 1986) based on the historians’ activities to decipher painters’ viewpoints represented as pictorial materials; the learning methodology has the following three phases: (1) **fact understanding**: grasp what each object depicted in historical cartoons symbolizes, (2) **relation understanding**: grasp what relationships among objects mean with giving careful consideration to the differences among them, and (3) **integrated understanding**: interpret the contexts of comprehensive historical backgrounds of the cartoons according to the painter’s viewpoints at that time by integrating the fact and relation understandings at the phases (1) and (2).

3. **Learning Activity Utilizing Semantically Enhanced Historical Cartoon**

Figure 1 represents an overview of our learning environment using semantically enhanced historical cartoons and its internal semantic representation. In order to let learners perform historical thinking activities, the learning environment embeds the three learning phases introduced in the previous section, namely (1) **fact understanding**, (2) **relation understanding**, and (3) **integrated understanding**. Learners perform respective historical thinking activities at each learning phase in a cross-sectional manner as necessary. In the system, in order to support their activities adaptively, learning phase and content dependent *inquiries* as stimuli are provided based on the semantics aware historical cartoons.

As far as the internal semantic information is concerned, we deal with the following four types of information: [a] contents of objects depicted in historical cartoons, [b] properties and relationships among objects, [c] artistic techniques called deformation used for strengthening painter’s messages in the historical cartoons and [d] gaze-aware transparent areas of respective objects in order to capture what learners are gazing at. In this research, we intend to capture the learner’s interests appropriately from learners’ gaze behaviors using eye-tracking devices based on the correspondence between the information of [a] and [d], i.e. the area of interest (AOI) regions on the objects in a historical cartoon. With respect to [c], Kuroda (1986) lays stress on the importance of artistic techniques illustrated in historical cartoons for deciphering the background contexts of the events. This research refers to the following five techniques (Historical Society of Pennsylvania): (i) **symbolism**, (ii) **irony**, (iii) **analogy**, (iv) **exaggeration**, and (v) **inference**. In order for the system to generate learner-adaptive inquiries at respective learning phases, we utilize a Question Generation Framework proposed by Jouault et al. (2016), in which inquiries can be generated by using the question templates prepared in advance based on the semantic information of [c] and other semantic information of [a], [b] and [d] for supporting learners’ historical thinking activities. Followings are the detailed explanations of each learning phase:

**(1) Fact understanding:** Transparent AOI regions are set on respective objects depicted in historical cartoons in advance to let the system recognize objects on which learners focus. In this phase, when the learner gazes at each AOI, the system dynamically generates inquiries (e.g., Q1 and Q2 in Fig.1) based on semantic information and a question generation template: e.g., on the object (i) in Fig.1, the semantic information ‘country is represented’, ‘coat is wore’ and the historical cartoon technique ‘symbolism’ are attached. According to the generated inquiries, learners try to grasp what each object symbolizes.

**(2) Relation understanding:** In this phase, the system infers the learner’s interests on relationships based on his/ her gaze behaviors among objects at phase (1), and gives inquiries to prompt his/ her historical thinking about them. Question 3 in Fig.1 is an example of the generated inquiry based on the semantic information about a relationship between two objects (e.g., on the relationship between (i) and (ii) in Fig.1, the semantic information ‘(i) aims for (ii)’ is attached as a type of [b] and the historical cartoon techniques, ‘irony’, ‘analogy’, ‘exaggeration’, and ‘inference’ are attached as a type of [c]). According to the inquiries, learners try to decipher painter’s historical viewpoints depicted as relationships between objects.

**(3) Integrated understanding:** In this phase, learners attempt to interpret the context of comprehensive historical backgrounds of the cartoons by integrating the fact and relation understandings at the phase (1) and (2), e.g., at the time when “The Sino-Japanese War” occurred, Japan’s ambition was to invade the Asian continent for expanding their power. Yoshikawa (2000) argues that the activities of micro-
level historical interpretation on backgrounds of each object depicted in historical cartoons contribute to enhancing macro-level historical interpretation activities. Although, it is difficult to provide content dependent inquiries adaptively to lead such interpretations due to the fact that learners’ interpretation varies according to their interests, the system generates inquiries about the objects to which the learner has NOT given attention at phase (1) and (2) (e.g., Q4 and Q5 in Fig.1). The generated inquiries serve as scaffolding to foster learners’ inclusive understanding activities for their fruitful interpretation activities.

![Figure 1. Overview of learning environment using semantically enhanced historical cartoons.](image)

**4. Conclusion and Future Works**

In this paper, we designed an adaptive learning support method that allows learners to perform their historical interpretation activities using semantically enhanced historical cartoons. In future work, we will endeavor to implement a learning support system based on the proposed design in this paper.

**References**


Developing a Geometric Proof Problem-Solving Support System Utilizing Card Selection

Megumi KURAYAMA
National Institute of Technology, Hakodate College, Japan
kurayama@hakodate-ct.ac.jp

Abstract: We think the solving of proof questions is cited as being conducive to the development of the skill of logical thinking. In this paper, we conducted an analysis of the structure of the proof question regarding the congruence of triangles. On the basis of this analysis, we constructed a solutions-support system that utilized the card selection method. As a result, the structure of the proof question consist of “assumptions” and “conclusions”. In this system, the blanks are then increased one after another, and finally, all the proofs are left blank. It is not only assumptions and conditions that are searched. In addition, sorting the given cards themselves and solving the proof promotes gradual learning.

Keywords: skill of logical thinking, Geometric Proof, Card selection, Problem-solving, assumptions and conclusions

1. Introduction

In recent years, the skill of logical thinking has received much attention. It can be said that this is because of an increase in situations that require expressing ideas logically, including interviews for jobs and for continuing on to higher education. Presentations and composition writing are often cited as being conducive to the development of the skill of logical thinking; however, besides these, the solving of proof questions can also be cited. Unlike presentations and composition writing, in the case of proof questions, things (i.e., assumptions or conditions) that one already knows about a given problem lead to conclusions that one wants to know, and the assumptions, conditions, and conclusions are already given. Therefore, this is believed to be easier for nurturing logical thinking when compared with other methods.

The question method of proof questions includes the “description method”, in which the learner describes everything from the beginning to the end, and the “fill-in-the-blank method”, in which a portion of the proof question is left blank and the learner describes the expressions or characters that correspond with the “blank.” However, when learners who use the fill-in-the-blank method suddenly utilize the description method to present a proof, it becomes necessary to think of the whole only from the part. Consequently, things that require thinking increase and give rise to difficulties. Therefore, it was considered possible to learn proofs through utilizing the “card selection method” proposed by Hirashima et al. (2012) and thinking of the structure from the part. Since, in the card selection method, it is essential to examine the cards prepared beforehand, the structure of the proof needs to be clear and it is vital to build a system. In this study, we conducted an analysis of the structure of the proof question regarding the congruence of triangles. On the basis of this analysis, we constructed a solutions-support system that utilized the card selection method.

2. The Structure of the Proof Question

Proof questions consist of “assumptions” and “conclusions,” and it is crucial to discover the conditions that lead to the conclusions and the elements essential to utilize these conditions and to solve them through description in accordance with an order. For example, the proof of the question in Figure 1 is shown as follows:
In the figure on the right, when points D and E, respectively, are taken as points on the line segments AC and AB; if AD = AE and ∠ADB = ∠AEC, prove that AB = AC.

Figure 1. Example of proof question

\[ \begin{align*}
\text{Proof} \\
\text{In } \triangle ADB \text{ and } \triangle AEC \\
\text{From the assumption } & \quad AD = AE \quad \ldots \ldots (1) \\
\text{From the assumption } & \quad \angle ADB = \angle AEC \quad \ldots \ldots (2) \\
\text{As the sides are common } & \quad \angle BAD = \angle CAE \quad \ldots \ldots (3) \\
\text{From } (1), (2), \text{ and } (3), \\
\text{as one side and the angles of the two sides, respectively, are equal, we get the following:} \\
\triangle ADB \cong \triangle AEC \\
\text{since, for congruent shapes, the corresponding sides are equal in length, we get the following:} \\
AB = AC \\
\end{align*} \]

(Q.E.D.)

When this proof is structured, it looks like Figure 2. However, the assumptions in the problem sentences have been omitted.

3. Problem-Solving Support System for Proofs

The screen of the proposed system is shown in Figure 3. Complete the proof question by moving the cards from the card group on the lower right-hand side to the blank part of the proof on the left-hand side using the “drag and drop” function. When all the blanks have been filled with the cards, you will be able to press the “answer” button. When the learner presses this button, the system determines whether the answers are correct or incorrect and provides the learner with an evaluation of right and wrong answers.

The card selection method refers to a method in which cards with simple sentences written on them are matched with blanks. In this method, as shown in Figure 3, it is only the part that can be understood from the problem sentences (including from assumptions and conditions) that is kept blank. Hence, the method is close to the fill-in-the-blank method and is not as difficult as the description method. It is easy even for beginners. Furthermore, as per Figure 4, if parts that form the sentence (such as conclusions or “from these”) are kept blank and cards that can be selected are increased, then, rather than just exploring the essential portion, one is required to think of the composition of the sentence. This results in a question format that is closer to the description method than it is to the fill-in-the-blank
method. At first, it is only the assumptions and conditions that are kept blank and simple problems are solved. As learners advance, they are first led to the conclusion that “the given figures are congruent,” and then the conclusions drawn are left blank. The blanks are then increased one after another, and finally, as shown in Figure 4, all the proofs are left blank. It is not only assumptions and conditions that are searched. In addition, sorting the given cards themselves and solving the proof promotes gradual learning.

4. Summary

In this study, we discussed the addition of blanks and cards in order to improve the proof problem-solving support system and the level of difficulty of the problems. Future challenges include the essential improvement of the system and having in the database information regarding the level of difficulty on the basis of the distribution map created from the results of the structural analysis of proof questions conducted in previous studies. Furthermore, it is important for problems corresponding to the level of difficulty to be created by gradually increasing and decreasing the blanks and the cards. Finally, it is essential to examine whether it is actually possible to advance learning through the use of the system.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 15K16109.

References

Analyzing Novice Programmers’ EEG Signals using Unsupervised Algorithms

Vanlalhuai SWANSA*, Tita HERRADURA** & Merlin Teodosia SUAREZa

aCenter for Empathic Human-Computer Interactions, College of Computer Studies, De La Salle University, Philippines
bCollege of Science and Computer Studies, De La Salle University – Dasmariñas, Philippines

*swansi@yahoo.com, trherradura@dlsud.edu.ph, merlin.suarez@delasalle.ph

Abstract: Ten (10) first year college programming students participated in the study and reported their emotions during the learning session. Emotiv EPOC headset was used to gather EEG brainwave signals. Digital signal processing filtering technique was used to filter the data. The reported academic emotions were engaged, confused, frustration and boredom. A square SOM map with 10 rows by 10 columns was built to visualize the EEG data set, a total of 100 nodes. The weights of the final SOM nodes were clustered using k-medoids and k-means algorithms, both derived two main clusters; one cluster aptly named “State of hope and enthusiasm” because it is primarily composed of clusters of confused emotion nodes surrounded by a topographical arrangement of engaged emotion nodes; the other cluster named “State of frustration and boredom” because it is primarily composed of frustrated and boredom emotion nodes. These observations of the topographical arrangements of the SOM nodes and its subsequent clustering of the SOM nodes by k-medoids and k-means, seem to be in accordance with previous findings by (Kort, Reilly & Picard, 2001; D’Mello & Graesser, 2011) ultimately making SOM to be a viable and good alternative representation/visualization tool for D’Mello’s theory of academic affect transition model. We also observed that k-medoids required much lesser number of k to derive similar clusters of SOM nodes as k-means, moreover, execution time for k-medoids is the same as k-means, making k-medoids a very attractive option for clustering algorithm of choice for clustering of SOM nodes.

Keywords: Affective Computing, Academic emotions, EEG data, Self-organizing Maps, Clustering algorithms.

1. Introduction

Students working on a complex task like solving problems, writing a computer program may experience varied emotions. Dr. Rosalind Picard stated that emotions play a vital role in learning since thinking and feeling are mutually present in normal human cognition (Picard, 1995). Thus, cognitive and affective states of a learner are crucial; similarly with affective transitions in order provide the necessary interventions to support learning (D’Mello & Graesser, 2012). Emotions are detected during learning by various physiological sensors (Frasson and Chalfoun, 2010). Electroencephalography (EEG) is a technique for reading scalp electrical activity (Teplan, 2002). Academic emotions are detected via EEG signals and are not easily faked (Mampusti, Ng, Quinto, Teng, Suarez and Trogo, 2011). The EEG signals are alpha, beta, delta and theta. Beta waves are high range frequency, an alert state, implying an increase in cognitive efforts (Boutros, Galderisi & Pogarell, 2011). This study deals with beta EEG brainwave signals.

(Craig et al., 2004) defined academic emotions as engagement being positive emotion and boredom, frustration and confusion as negative emotions and should be handled in an Intelligent Tutoring System (ITS) for necessary intervention to support learning.
2. Related Works

Transitions occur between academic emotions (D’Mello & Graesser, 2011), when progress is blocked, a student experiences confusion, when resolved, it transitions to engagement (see Figure 1). When confusion is not resolved, frustration occurs resulting in interplay between confusion and frustration. When frustration persists, boredom sets in, resulting in interplay between frustration and boredom (Bosch & D’Mello, 2013; Bosch & D’Mello, 2015).

3. Methodology

The participants were ten (10) first year college students from DLSU–Dasmariñas, Philippines. Raw EEG signals were collected from students engaged in computer programming using Emotiv EPOC head set. Data gathering techniques were based on the work by (Azcarraga, Marcos & Suarez, 2014).

Beta waves were extracted using a digital bandpass filter (1000th-order) with passband of 14-30 Hz. The bandpass filter used hamming windowing method. The filtered signals are then subjected to Fourier analysis and statistical features were extracted from the dataset.

4. Results and Discussions

The EEG data set with 48 features and 4,000 instances were analyzed using unsupervised learning algorithms like Self Organizing Maps (SOM), k-medoids and k-means clustering algorithms.

(Kangas, Kohonen and Laaksonen, 1990; Ritter and Kohonen, 1989) used SOM to represent abstract data relationship via topographic maps. A square map with a total of 100 nodes was built. Initial neighborhood was the entire SOM map and highest initial learning rate was 0.9 and 0.1 the lowest. To ensure fair selection process a random single sample x was selected from the dataset and exponential decay function was used. SOM was implemented two times, implementation #1 seems to have stabilized at 35,000th iteration and implementation #2 seems to have stabilized at 30,000th iteration.
To identify similar nodes of the SOM, K-Medoids and K-Means clustering algorithms were applied. K-medoids takes the most centrally located object in the cluster (i.e., median) and K-means takes the mean value (i.e., average). Figure 2 shows a SOM with K-medoids when k=7 and Figure 3 shows a SOM with K-means when k=16, the clusters derived by both algorithms are very similar to each other.

The right hand side of the map in Figure 2 and Figure 3 is dominated by confused and engaged emotions and appear to conform to the state of hope and enthusiasm as mentioned in (Kort, Reilly and Picard, 2001; Rodrigo, Baker and Nabos, 2010). The left hand side clusters of the map is dominated by frustrated and bored emotion nodes and labeled as state of frustration and boredom and appear to be consistent with (Graesser and D’Mello, 2011)’s cognitive disequilibrium model which states that confusion has to be resolved and if it goes unresolved, confusion will lead to frustration and boredom.

K-medoids required much lesser number of k (k=7) to derive similar clusters of SOM nodes as K-means (k=16) and execution time for k-medoids is the same as k-means, making K-medoids a very attractive option for clustering algorithm of choice for clustering of SOM nodes.

5. Conclusions and Recommendation

SOM allowed us to determine relationships based on which nodes are adjacent to each other while k-medoids and k-means determined the relationship based on which nodes are included in the cluster, making the analysis more meaningful and interesting. We may infer that SOM is a viable and good alternative representation/visualization tool for D’Mello’s theory of academic affect transition model.

Future works include increasing the number of participants and implementing the above analysis on adult learners’ EEG dataset.

References


Evaluation of Mathematics Knowledge Level through Personalized Learning Exercise based on the Adaptive Tests

Yan LIU, Chuxin FU & Xiaoqing GU*
East China Normal University, Department of Educational Technology, Shanghai, China
* xqgu@ses.ecnu.edu.cn

Abstract: Personalized exercise pursues the goal of the most effective improvement by least exercise. How to recommend exercises for students to achieve this goal is the core and key issues of personalized exercise. This paper analyzed the feasibility of applying adaptive testing in personalized exercise, and adopted non-equivalent control group pre-test post-test designing. The result has shown that applying adaptive test can effectively improve the students' ability level and performance.

Keywords: personalized learning exercise, Adaptive Testing, IRT, Feedback Theory

1. Introduction

There were so much homework for students, which probably make repeat exercise in vain in current K-12 schools. According to knowledge gained level, it is necessary to acquire students’ skills to organize personalized learning exercise. Lord’s (1980) theory was applied that if a question, which was neither difficult nor easy for the students, would be a good question. In other words, the core of personalized learning exercise is finding out appropriate questions.

Item Response Theory (IRT) (Edelen & Reeve, 2007) was a new measure theory breaking through classic measure theory limitations, firstly known as the potential trait theory. The most important feature of IRT compared with classic theory is the invariance of both item parameters and ability parameters, and two parameters are placed in the same scale. Learning Feedback was proposed for the first time in 1948 by the founder of cybernetics (Wiener, 2000). In the field of education, feedback can help learners to modify their own ideas and behaviors to promote learning (Bangert, et al., 1992). Moreover, it also can help teachers to adjust teaching plan to adapt to each levels of learners (Shute, 2008).

2. Methodology

2.1 Participants

The study selected 44 students from two Grade 8 classes taught by the same math teacher in a Shanghai Junior Middle School. They were divided by the experimental group, and the control group, the experimental group was adopted adaptive test.

2.2 Hypotheses

The first hypothesis was that the average increase in the number of students using the individual exercises would be better than the student using the general exercise. The second hypothesis was considered that the average increase of tests using personalized exercises would be higher than the average increase in scores for tests using general exercises.
2.3 IRT Algorithm

This study mainly discussed logistic model of 0-1 scoring single-dimensional IRT model using three-parameter Logistic model (3PLM)( Birnbaum,1968), the formula is as follows:

\[ P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-D(a_i(\theta - b_i))}} \]

i is the i-th item; D is a constant,1.7; \( a_i \) is the degree of discrimination for the i-th item; \( b_i \) is the difficulty of the i-th item; \( c_i \) is the guess coefficient of the i-th item; \( \theta \) is students’ ability level estimated according to test performance.

2.4 Experiment Processing

First, we used R to calculate theta and create items. Second, the study has assigned learning exercise tasks as shown in Figure 1, the experimental group was adopted adaptive test while control group did not.

3. Results and Discussion

There were four exercises conducted for two groups. After each exercises completing, theta values were estimated based on the students’ answering results, respectively recorded as theta 1, theta 2, theta 3, and theta 4. The initial theta value was recorded as theta 0. Firstly, the study was used ggplot2 package of R language to draw theta value of the probability density function diagrams (Figure 2), and theta value distribution was similar with the normal distribution.
First, mean of $\theta$ of the Experiment group has changed, however, change of means among Control group was not obvious; the standard deviation of the Experiment group and the Control group varied a little. The minimum value of $\theta$ of the Experiment group has improved quite a lot (theta0=-1.238, theta4=-0.745), however, the control group was not obvious (theta0=-2.540, theta4=-2.466).

3.2 Overall Change of Students’ Achievement

In order to verify the effect of practice intervention on student achievement, teacher sent the previous unit test (pre-test), which were with equivalent difficulty. Independent sample T test was conducted to exam the pre-test between experimental and control group. The result indicated that the variance of the two groups before the test results are missing ($F = 7.957, p < 0.05$) in the Table 1, which has been to deny the variance equivalent hypothesis. The result of variances, which were not equal ($t = -3.806, p < 0.05$), have showed that pre-test scores was significantly different in different groups. The results showed that personalized learning exercise can effectively improve the student's theta value and scores.

Table 1 Independent sample t-test results of pre-test scores

<table>
<thead>
<tr>
<th></th>
<th>Levene test of variance equation</th>
<th>The t-test of mean equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>Sig.</td>
</tr>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>variance are equal</td>
<td>7.957</td>
<td>.007</td>
</tr>
<tr>
<td>variance are not equal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusion and Limitations

The concept of personalized learning exercise has been defined that we would apply personalized feedback to student by adaptive test to diagnose each test results. The personalized training recommended mechanism to achieve the goal is the key to solve problems. The results showed that: (1) In general, when the amount of items is not handful, random exercises has been no significant effect on students' ability level. (2) In terms of student outcomes, the individualized learning exercise has been improved the value of theta and scores in each test. In the future research, the number of participant should be increased. The future study will consider various impacts, such as learning strategies, to modify the adaptive test arithmetic.

References


Discovering Dynamics of an Idea Pipeline: Understanding Idea Development within a Knowledge Building Discourse

Alwyn Vwen Yen LEE & Seng Chee TAN*
Nanyang Technological University, Singapore
*sengchee.tan@ntu.edu.sg

Abstract: Idea development is an important process within a knowledge-building discourse and it is crucial to understand the dynamics of idea development throughout the discourse, such as the growth, flourishing or fading of ideas. This study proposes a framework called Idea Pipeline that explores and tracks the dynamics of idea development within a knowledge-building discourse. This pipeline consists of three phases: discovery, identification and analysis, and ‘rise above’ of ideas. Each phase of the pipeline will be illustrated using findings from a comparison study of two online knowledge building discourses. During the discovery phase, a text miner is used to identify groups of related keywords from the discourse; this is represented as keyword graphs with weighted frequencies to show the diversity of ideas that were embedded within the knowledge-building discourse. In the idea identification and analysis phase, network analysis was conducted to help label key ideas that were promising to the discourse community; this would provide the community with information to decide which ideas to pursue so that advancement of communal knowledge could be achieved leading to the ‘rise above’ phase. This Idea Pipeline framework can be an additional method for the temporal analysis of a computer-supported collaborative learning discourse over a longer duration of weeks or even months.

Keywords: Online discourse analysis, knowledge building discourse, idea pipeline, idea development, text mining, network analysis

1. Introduction

In the field of computer-supported collaborative learning (CSCL), researchers have access to processual data over a duration of weeks, months, or even years. Yet, many CSCL researchers ignore the temporal information and choose not to analyze changes over time. Reimann (2009) argued that experimental studies that have strict control for variables may not work well for CSCL studies, not just because of the practical difficulties in controlling the variables, but also the fundamental assumption that the factors exerting consistent and continuous influence on the learning events over the entire period may not be valid. Few CSCL researchers have studied processual changes over time, and among the few (e.g., uptake of ideas by Suthers, Dwyer, Medina, & Vatrapu, 2010), the current methods rely extensively on labor-intensive manual coding and interpretations. In addition, methods such as a conversational analysis that examines the temporal development of ideas are not applicable in explaining longer term unfolding of events.

This paper intends to address this research gap in CSCL by proposing a framework called Idea Pipeline that explores and tracks the dynamics of idea development within a knowledge-building discourse over time. We piloted and illustrated the workings of this framework using a comparison study of existing data, which were obtained from two online knowledge building (KB) discourses conducted on Knowledge Forum (Scardamalia, 2004). We showed how technologies, such as text mining and network analysis tools, can be leveraged to provide insights into learning behaviours in online discourse and to automate part of the analyses. The analyses in this study were carried out after the completion of a course, but the methods can be applied to facilitate knowledge building among learners at various junctures of their discourse.
2. Background

As knowledge is constructed through social interactions such as discourse (Stahl, 2004), researchers have been harnessing technological tools to investigate discourse, such as converting knowledge and ideas into innovative product (Hansen & Birkinshaw, 2007), or assisting students in making promisingness judgements to achieve better knowledge advances that benefit their community (Chen, Scardamalia & Bereiter, 2015). This study is situated in knowledge building classrooms (Scardamalia & Bereiter, 2015), where collaborative improvement of ideas forms the central tenet of students’ learning. Online discourse is conducted on Knowledge Forum, a CSCL platform that is specifically developed to support knowledge building using discourse scaffolds and analytics. Knowledge building engages students in discursive activities intended to enhance collective understanding (Bereiter, 2002). Students in a knowledge building community work on authentic knowledge problems that arise from their attempts to understand the world, and the knowledge problems are often identified and articulated as ideas. Students construct and progressively improve on ideas represented as knowledge artifacts (such as online notes containing their ideas) and this process often leads to a diversity of competing ideas. This provides the natural impetus for discussions that lead to collective idea improvement. By improvement of ideas, we refer to the enhancement of coherence, quality, and utility of ideas. Continual engagement in idea improvement process may lead to ‘rise above’, which refers to the higher-level outcome of integrating several ideas or framing the inquiry using a higher-level principle or theory.

The process where students deliberate, work on and improve their shared ideas collaboratively is similar to the innovation processes of improving the design of an object, such as an improved version of a mobile phone. The classical view of the innovation process consists of four phases: idea generation, idea formulation, problem-solving, and utilization (Myers & Marquis, 1969). In the innovation process, a large number of ideas are generated in the early stages, which need to be evaluated and selected, so that resources can be devoted to more valuable and promising ideas, which could culminate in innovative products or solutions that are critical to the organizations. This process has been largely implemented in production lines and large organizations such as Microsoft Corporation (Bailey & Horvitz, 2010). Nevertheless, similar processes of discovery, identification, analysis and application of ideas can be adapted for a learning community to engage learners in deep learning and meaning making, while steering the students away from focusing solely on knowledge acquisition.

The introduction of the Idea Pipeline framework is significant in several ways. First, in terms of methods, digital data associated with the online discourse allows for some degree of automation, such as the use of analytical tools to calculate and measure learning outcomes, and to identify and analyze ideas within discourse (Lee, Tan & Chee, 2016). In other words, technologies such as text mining and learning analytics can be leveraged to automate parts of the analysis of student interactions and discourse. Second, the proposed Idea Pipeline can track and monitor the development and movement of ideas within the learning community, while seeking to understand the dynamics that stimulate the growth, flourishing or fading of ideas within discourse. This framework can be useful to the CSCL community, as the process of analyzing ideas in discourse over time is complex and still requires substantial research in determining development trajectories and interactions of ideas among different agents in the community. The Idea Pipeline allows us to form a deeper understanding and comprehends idea developments visually within a knowledge building discourse in a more succinct manner. The methods proposed in this study are also different from Big Ideas Tool (Chen, Chuy, Resendes & Scardamalia, 2010) that requires the students to identify ideas they feel are promising.

3. Idea Pipeline – A Framework for Tracking and Analyzing online Discourse

The proposed Idea Pipeline framework consists of three sequential phases (Figure 1), with emphasis placed on the processes of evolution of ideas through phases of (a) discovery, (b) identification and analysis, and (c) ‘rise above’ of ideas. First, a pool of ideas is discovered from a communal discourse using a text mining tool to determine related keywords from the discourse with weighted frequencies to show the diversity of ideas and relationship among ideas embedded within the knowledge-building discourse; this can be visualized as keyword graphs (see Figure 2). As most ideas within the initial
phase are competing and often in preliminary forms with uncertain prospects, identifying promising ideas is the key focus in the second phase of Identification and Analysis. Ideas that can bring the community forward to help achieve a higher level of understanding are considered promising ideas (Chen et al., 2015). Promisingness can be considered from either the student’s or the teacher’s perspectives. For example, to a teacher, a promising idea is aligned to a learning objective. To a group of students, a promising idea has potential to lead to their shared learning goal. This study analyzes promisingness of ideas from the students’ perspectives based on three criteria: (a) the relevancy of ideas to the context; (b) the level of interests of the community in pursuing ideas deserving of further development; and (c) the degree of impact that ideas have on communal discourse through synthesis and improvement of ideas. Finally, the ‘rise above’ phase in this study consists of integrating diverse ideas to achieve a higher level of understanding, often manifested as a deeper conceptual representation, or higher level of abstraction by using a theory or principle in the explanation for an inquiry. For students who were not able to achieve ‘rise above’ through communal discourse, results from the first two phases of the pipeline can inform the students about other promising ideas present in the discourse, and they can choose to reestablish their interests and continue contributing to the discussion.

![Figure 1. Idea Pipeline Framework](image)

In terms of technology and methods, a text mining tool called SOBEK (Reategui, Epstein, Lorenzatti & Klemann, 2011) is used in the first phase of idea discovery. SOBEK can identify the ideas (as keywords) in a text and construct a relational keyword graphs. Second, building on a previous study (Lee et al., 2016), the Idea Identification and Analysis methodology (I²A) is applied in the second phase of the Idea Pipeline. In essence, the I²A methodology identifies and evaluates a pool of ideas within a discourse by using a combination of network analysis and temporal analytics to identify promising ideas that relate to context and are significantly interesting to the community. The usage of network analysis helps to assess the interactivity between different agents and components of the discourse network, while temporal analytics provide insights into the often-overlooked temporal dimension of discourse data. The third phase of the pipeline describes the ‘rise above’ of ideas through qualitative textual analysis. This ‘rise above’ phase is complex, usually covers a wide range of developments, and is represented mostly through artifact developments, changes in learning approaches or behavioral changes. It can include a translation of verbalized intentions into actual action or practice that are often not captured through discourse. As automation of the ‘rise above’ phase was not a goal for this study; instead, a qualitative approach was adopted for the analysis of the ‘rise above’ phase.

4. Knowledge Building Classes – Context of the Study

The two knowledge-building classes in this study were two cohorts of educators enrolled in a graduate-level course in two different years, but the courses were conducted using similar content and resources. The course focused on understanding CSCL and knowledge building by encouraging the participants to act as knowledge builders. Community A consisted of 15 participants, including two lecturers and 13 graduate students. Community B consisted of 11 participants, including two lecturers and nine graduate students. This study analyzed the textual discourses (Discourse A and Discourse B) and the dynamics related to the idea development process. Both communities underwent 13 weeks of instruction and collaborative discourse on topics of CSCL and knowledge building as part of their
master’s in the education program. Knowledge Forum was used as the main online discourse platform for students to communicate and share ideas with one another. The teaching staff helped to facilitate learning and co-creation of knowledge within the knowledge building community. The notes on Knowledge Forum, referred to as ‘discourse units,’ were periodically archived for reference and was subsequently anonymized for further analysis at the end of the course.

Over 13 weeks of 3-hour sessions, the participants discussed about CSCL and the core principles of knowledge building, the ways of designing and facilitating collaborative improvement of ideas, and how they could apply what they have learned to design learning environments for their students. This study focuses on investigating how the dynamics of idea development – including processes such as generation and discovery of ideas, sharing of diverse ideas and improvement of such ideas – are manifested in a knowledge building discourse (Scardamalia, 2002). As we illustrate the working of the Idea Pipeline framework, we seek to answer the following research questions: (1) What are the key ideas pursued by the two classes of graduate students in their respective discourses? (2) What are the promising and potential ideas that can be identified from their discourses? (3) Was there any evidence of higher level of understanding (‘rise above’) found in their discourses?

5. Methods

5.1 Text Miner SOBEK

In this study, the key ideas in the discourses were harvested from the participants’ notes in the Knowledge Forum. SOBEK (Reategui et al., 2011) was used to identify relevant keywords and related concepts from unstructured text data, using frequency analysis of textual material and filtering of important keywords. SOBEK generates relational graphs of keywords – a graphical representation of nodes of keywords with relative sizes to indicate their frequencies and connections to indicate relationships among the keywords. An in-built thesaurus within SOBEK helps to minimize unnecessary repetition of keywords by filtering out common words such as related prepositions and adverbs, and further helps in aggregating words belonging to similar concepts and meaning.

5.2 Idea Identification and Analysis (I2A) methodology

I2A (Lee et al., 2016) is a mixed method approach that uses network analysis with keywords and temporal analysis of textual discourse, to identify and trace the development of ideas that become promising talking points within a discourse. We first used a network measure, the betweenness centrality (BC) coefficient, to assess meaningful connections between keywords within discourse units (notes in Knowledge Forum). The presence of keywords in a discourse unit represents the key theme or meaning of ideas within the discourse unit. When the BC values of discourse units are calculated over time, the BC value of each discourse unit will vary according to the network changes and the presence of keywords in the discourse unit; the plot of BC values over time forms a BC temporal trend for each discourse unit. The pattern of the BC trends indicates the nature of the ideas contained in a discourse unit. For example, a discourse unit showing a trend of sustained BC values over time is likely to contain communally interesting and promising ideas. Qualitative analysis is used to validate this classification result. In the previous study (Lee et al., 2016), keywords were provided by the teacher, and these keywords were used for identifying promising ideas to the teachers. In this study, I2A was deployed similarly, but the keywords were extracted from the student discourses, and therefore, it was used to determine ideas promising to the students.

The promisingness of an idea, when assessed from the student’s perspective, is based on three criteria: (a) the relevancy of ideas to context; (b) the level of interests of the community to pursue ideas deserving of further development; and (c) the degree of impact that ideas have on communal discourse through synthesis and improvement of ideas. When ideas in the discourse satisfy these three criteria, the ideas are considered as promising ideas. These are ideas that the community feel are relevant and are interesting to continue pursuing after some initial interest. More time can be allocated for discussion, such that interests in these promising ideas can be sustained within the knowledge building discourse. The community can then build on each other’s ideas, achieve ‘rise above’ as a community and
eventually achieve a higher level of understanding and advancement of communal knowledge. However, the proportion of promising ideas in discourse is often far and few. It would be useful to apply less strict criteria to uncover ideas with a lower degree of relevancy, communal interests and impact. These ideas are labelled as potential ideas that may contain less promising but yet important elements that help advance discussions.

5.3 Finding ‘Rise Above’ from Discourse

The method to identify ‘rise above’ of ideas in a discourse is difficult due to its wide range of developments that can occur simultaneously throughout the whole discourse. Although ‘rise above’ represented through artifact developments are easier to track using quantitative methods (e.g., finding specific phrases or usage of scaffolds), the changes in learning approaches and behavioral changes are, however, harder to find. Therefore, a qualitative approach was adopted in this study to find discourse units with ‘rise above.’ We first used the I2A methodology to search for discourse units that contain promising ideas, and scrutinized the contents for any ‘rise above.’ We subsequently broadened our search to look for ‘rise above’ in discourse units that are related to the identified discourse units. Notably, ‘rise above’ notes contained shared experiences from participants and reflection of their follow-up actions towards some group activities or individual tasks. The experiences and actions are often influenced by feedback from the community or are drawn from inspiration that originated from promising ideas found within the discourse. ‘Rise above’ notes can also be authored by multiple authors and be the result of group discussions conducted over several group meetings.

6. Findings and Discussions

6.1 Discovery Phase

We ran the two knowledge building discourses (281 notes in Discourse A; 204 notes in Discourse B) through SOBEK and obtained a table of weighted frequencies and two relational keyword graphs that contain relevant and commonly used keywords within the respective discourse. Results in Table 1 show that Discourse B contained fewer unique keywords as compared to Discourse A. The frequencies of keywords in Discourse B were adjusted with a proportional weighting to account for the different number of notes, and to further provide a fairer comparison of keyword frequencies between the two discourses. The weighting shown in Table 1 is based on the difference in note count (281/204=1.38x), with the assumption that all notes have equal possibility of contributing promising content to the discourse.

<table>
<thead>
<tr>
<th>Keywords between the two discourses</th>
<th>Discourse A, 1x weight (Original frequency)</th>
<th>Discourse B, 1.38x weight (Original frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based</td>
<td>63 (63)</td>
<td>No significant presence</td>
</tr>
<tr>
<td>Community</td>
<td>79 (79)</td>
<td>No significant presence</td>
</tr>
<tr>
<td>Discourse</td>
<td>82 (82)</td>
<td>No significant presence</td>
</tr>
<tr>
<td>Idea</td>
<td>67 (67)</td>
<td>118.68 (86)</td>
</tr>
<tr>
<td>Information</td>
<td>55 (55)</td>
<td>No significant presence</td>
</tr>
<tr>
<td>Knowledge</td>
<td>325 (325)</td>
<td>248.4 (180)</td>
</tr>
<tr>
<td>knowledge building / kb</td>
<td>424 (424)</td>
<td>313.26 (227)</td>
</tr>
<tr>
<td>Learning</td>
<td>172 (172)</td>
<td>238.74 (173)</td>
</tr>
<tr>
<td>Process</td>
<td>No significant presence</td>
<td>75.9 (55)</td>
</tr>
<tr>
<td>Students</td>
<td>151 (151)</td>
<td>193.2 (140)</td>
</tr>
<tr>
<td>understanding</td>
<td>114 (114)</td>
<td>81.42 (59)</td>
</tr>
</tbody>
</table>

In this study, both communities of students underwent similar knowledge building activities. The ideas discovered within the discourses are a reflection of authentic thoughts, opinions, and actions that participants contemplated. This situation was observed to be consistent throughout both discourses.
as both communities sought to build new knowledge and apply their knowledge on top of their current teaching roles in schools. The results show that Discourses A and B had 10 and 7 identified keywords respectively, with Discourse A showing a more diverse range of ideas. The larger number of keywords and connections between keywords in Discourse A showed that Community A, with the relatively larger number of participants and notes, generated a more diverse list of ideas.

Further, even though both communities contain six common keywords (see Table 1), the differences in the discussion foci and content for both communities are revealed in the relational keyword graphs. The relational keyword graphs were generated shortly after the discourses started (Figure 2) and after the discourses ended (Figure 3). We collapsed the term ‘kb’ with the keyword ‘knowledge building’ since both keywords refer to the same concept. The keyword graphs show a stark contrast in the development of ideas in the respective discourses. One significant difference is in the starting points of discourse for both communities. Community A began the discussion on knowledge building, whereas Community B started to discuss on the topic of CSCL, even though both communities were provided with the same teaching materials and instructed by the same teaching staff. Over the period of discourse, the relational keyword graphs for both communities expanded but Community A was more concerned with exploring the concepts of ‘knowledge building,’ ‘community knowledge’ and ‘understanding,’ while Community B was more interested in discussing ‘student learning’ and the ‘learning process.’

![Figure 2. Different discussion foci for Discourse A (left) and Discourse B (right) at the initial part of the discourse](image)

![Figure 3. A more diverse spread of keywords in Discourse A (left) than in Discourse B (right) at the end of the discourse](image)

As knowledge building is the foci in both courses, both keyword graphs show that ideas related to the conceptions of ‘knowledge building’ have been communicated and discussed between students in their respective discourses. The keyword network of Discourse A has noticeably higher connectivity, repetition, and a variety of keywords that were used to represent a wider spread of ideas, opinions, and thoughts. Burchfield (1996) proposed that increasing repetition of keywords that are important to the reader’s comprehension allows discourse participants to readily connect ideas and increase coherency. One can also argue that this is plausible at the superficial level, but the issue of how knowledge coherence is achieved is not trivial. This study does not depend on keyword frequency alone to assess cohesiveness, but rather, combines the visualizations of keyword frequencies with inter-keyword connections to show the intensity of keyword usage and strength of connections between keywords in the discourse. Thus, Discourse A is likely to contain arguments that are more cohesive. There were fewer edges in Discourse B (connections between different nodes in the network) observed in the keyword graph, which is a reflection that the ideas were not properly connected in the discourse. The
presence of isolated keywords or keywords not well connected to other related keywords shows the lack of coherence among the ideas in Discourse B. This might explain the lack of activity and conversations in Discourse B, as compared to the wider range of activities and discussions found in Discourse A.

In relation to the content of discourse, a more detailed scrutiny of the textual discourse provided insights into the intentions and learning objectives set out by both communities. Despite the fact that both communities were using the same materials taught by the same teaching staff, with the same instructional objectives, other than the list of six common keywords (Table 1) that are related to the central theme of knowledge building, the overall foci for both communities were found to be quite different. Community A placed greater emphasis on understanding the concept of using knowledge building as a form of enhancing communal understanding, whereas Community B gravitated towards application – using knowledge building and technological tools to help students improve their individual learning processes. In summary, from the analysis of Discourse A, we discovered that knowledge building discourse among the participants did help in the contribution of diverse ideas. At the same time, even though the presence of diverse ideas and unique viewpoints serve as inspiration for other community participants, there remains a need for the community to eventually come together and take responsibility for advancing group knowledge. With regard to Discourse B, although the initial topic of CSCL dominated early parts of the discourse, the lecturer guided the community back to discussions regarding knowledge building by the end of discourse. Discourse B was able to reach a consensus in due course by the end of discourse, albeit with a lower diversity of ideas.

6.2 Identification and Analysis (I²A) Phase

In essence, I²A methodology can be used to identify promising ideas through network analysis and measures such as betweenness centrality. Groups of keywords discovered using SOBEK in the discovery phase were used to calculate BC values, which are indicative of the importance of keywords and centrality within the discourse network. When BC is calculated over time, BC temporal trends are formed, and these trends constitute patterns that can be used for classifying ideas within discourse into different idea types, such as promising and potential ideas. For discourse units with BC trends that exceeded the BC thresholds, the discourse units are considered to possibly contain promising ideas. In this study, the BC values that act as thresholds for promising ideas are calculated to be close to 0.026 and 0.049 for Discourse A and Discourse B respectively, and the thresholds for potential ideas are calculated to be slightly lesser (<10% of BC threshold for promising ideas) for potential ideas. The final number of promising and potential ideas for each discourse are shown in the results (Figure 4).

![Figure 4. Frequency of promising and potential ideas for Discourse A and Discourse B](image)

In a straight-up comparison between Discourse A and B, Discourse A contains more promising and potential ideas, indicating high relevancy of ideas to context and extensive interests from the community. The results show that there were six promising ideas in discourse A and nine potential ideas, which can be further improved to become promising talking points. The analysis in the earlier idea discovery phase has shown that there is a larger diversity of keywords and ideas in Discourse A than Discourse B. The results obtained from the I²A analysis provide corroborative evidence that cohesive arguments and diversity of ideas in Discourse A have provided more interesting talking points that the community can participate in. Eventually, the increased interests and participation in discourse allowed the community to attain a higher level of understanding through discussions. Discourse B contained fewer keywords and diversity of ideas, and there was a corresponding fewer promising and
potential ideas. This suggests that it is less likely to find significant interests or participation from participants in Community B, and ideas in Discourse B are likely to have a muted impact on subsequent parts of the discourse.

Since each discourse unit is an individual node connected with others in the discourse network, each discourse unit has its BC trend, and there would be the same number of BC trends as the total number of discourse units in a discourse. Overall, the plots of most BC trends tend to increase to a peak, then decreases over time, and is consistent with the temporal analysis of BC trends using network analysis. However, a graph containing BC trends of all discourse units can appear disorienting and mask the key trends of Discourse A and B. Further, we are more interested in analyzing the unique behaviors of BC trends that vary over the period of discourse, such as the significant or abrupt changes in BC values at specific temporal positions of the discourse. To better reflect the major difference between Discourse A and B, in this study, we condensed and averaged the majority of significant BC trends and their values within a particular discourse into a single BC trend. This BC trend is labelled as a generalized BC trend that is representative of most of the BC trends in either Discourse A or Discourse B. The overall general shape of BC trends for both discourses were plotted (Figure 5) and critical differences can be identified from the two trends. The generalized BC trend of Discourse B exhibited an abnormally high peak early in the discourse, but it dropped to insignificant levels soon after. Discourse A had a gradually increasing generalized BC trend with local peaks towards the middle of the discourse, before decreasing and plateauing at a slightly lower than average BC value towards the end of discourse. Both BC trends were truncated at discourse frame 70, as there were no significant BC trend movements observed after that.

![Figure 5. Generalized BC trends for Discourse A (solid line) and Discourse B (dotted line)](image)

Further examination of the textual data provided corroborative evidence consistent with the earlier observations and analysis from the discovery phase of the Idea Pipeline. Participants in Community B initially embarked on discussions on CSCL, but they were eventually guided back to the topic on KB, and the cooling of interests in discussions after the switch is reflected by the significant decrease in BC trend. On the contrary, Community A (solid line) embraced the topic of KB right from the start of discourse and subsequently engaged in discussions to further improve ideas and ‘rise above’ through collaborative knowledge building discourse. There was heightened interest among the participants, reflected by the peak in the BC trend, before communal consensus brought the discussions to a close with lower activity towards the end of Discourse A. Overall, Discourse A contains sufficiently interesting promising ideas that became talking points, which sustained the interests of the community to continuously build on and improve each other’s ideas in order to achieve better understanding. Discourse B was not able to generate the same amount of communal interests, and results indicated that discussions were also not as broad and extensive. These results show that the usage of I2A within the Idea Pipeline allows us to identify and analyze the flow of ideas and interests within the community, in a more visible and scalable manner.

6.3 ‘Rise Above’ Phase

This phase was difficult to analyze using quantitative measures or with the aid of analytical tools, as the extent of ‘rise above’ of ideas by participants cannot be accurately measured or tracked when actions and consequences occur outside of online discourse. However, we were able to use the I2A
methodology to identify discourse units that contain promising ideas, and the methodology helped us save time in searching for ‘rise above’ in the discourse. The search was expanded to look for a qualitative interpretation of instances, where shared experiences from participants and consensus from the community helped sustained interest in discussions that eventually improved and advanced communal knowledge. These instances of discourse provided insights into the extent and success of discussing and integrating ideas that were deemed promising by the community, and are quite frequently spotted within Discourse A. We extracted two examples from Discourse A to illustrate how participants built on promising ideas, improved the ideas throughout the course, and integrated peers’ ideas to achieve deeper understanding of conception of knowledge building.

Let me quote my example: Initially, I have the problem of understanding whether kb is a theory, pedagogy or technology. Throughout this course, I have read our classmates' different perspectives, ideas pertaining to this problem of understanding and I tried my very hard to synthesize these ideas and arrive at my theory that kb is a theory … This synthesis of idea leads me to ask more questions such as is kb related to the theory of connectivism? Can other alternatives technologies be used to support a principle-based pedagogy? Do you think I am achieving rise above? or which dimension of rise above do you think I belong? (DU268 by student S6)

I came into this course seeking a validation - that I am already practising kb and that kb is pbl (problem-based learning), pbl is kb. As I read more on kb and as I prepared the discussion on pbl & kb, I realised I have obtained my rise above. I now have a much clearer idea to this problem of understanding on ‘what is the difference between kb and pbl?’. I have learned that pbl and kb while highly similar has distinct differences, and I should not be attempting to merge the pedagogies as one, though they are very similar … Understanding a pedagogy would be trying to understand its paradigm of education and understanding its origin. (DU269 by student S11)

Both students S6 and S11 mentioned that they achieved a better understanding of their own inquiries and problems through discussions and self-directed learning by the end of the discourse. They discovered and identified conflicting ideas that were viewed as a problem, and analyzed it through discourse with the community, in the form of sharing sessions and presentations. The students eventually worked on their ideas with further intention to improve ideas, either through posing of questions by student S6 to invoke responses or through self-reflection evident in student S11’s case, in the hope of ascertaining that his/her statements and contributions are accepted by the community. By undergoing such a process of idea development through the Idea Pipeline, novel ideas can be discovered with a new perspective and be improved through sharing over multiple idea development cycles, such that eventually, the community as a whole can achieve a better level of understanding through communal discourse.

There was, however, no significant ‘rise above’ of ideas in Discourse B and it could be due to a few reasons. First, the initial deviation of discourse topic required a concerted effort from both the community and teaching staff to get the community back on-track, and it could have detracted the community from focusing on topics that they have interests in. From another perspective, the sentiment within Community B is often muted and less enthusiastic as compared to Community A, therefore pointing to a possibility that Community B might have little genuine interest in discussing the topics.

7. Conclusions and Future Directions

In summary, we proposed the usage of the Idea Pipeline framework to discover the dynamics of the idea development process, by tracking and monitoring ideas, through the phases of discovery, identification and analysis, and ‘rise above’ of ideas in a discourse. By using a comparison study of two knowledge building discourses, we illustrated how promising ideas can be identified and how these ideas improve and evolve in communities of different learning approaches and engagement. The comparison studies demonstrated how the Idea Pipeline and associated methods can be used, with different types of communities and discourse, to analyze textual data for visualization of idea diversity and identification
of promising ideas in knowledge building discourse using network analysis. The identification of promising ideas in discourse is critical to knowledge building, as communal interests need to be sustained for the process of continuous improvement of ideas, so that ‘rise above’ can occur to benefit the learning community. The Idea Pipeline framework and proposed methods provide a way to uncover this critical information to the community, through discourse analysis, network analysis and usage of temporal analytics. In our pilot study, by observing information such as the visualization of scarce keywords in the keyword network and the lackluster BC trend during mid-discourse, teacher interventions could be introduced to help Community B eventually achieve more ‘rise above’ by the end of the knowledge building discourse. Moving forward, we intend to study how a knowledge building community can use the Idea Pipeline framework and methods for near real-time monitoring of idea creation and development processes so that the participants could be guided towards a more productive knowledge-building discourse. We have also proposed the Idea Pipeline framework to schools and teachers with the aim of studying larger population to validate and enhance the Idea Pipeline framework. In addition, our future work include development of a systematic method to identify and characterize instances of ‘rise above’ of ideas in knowledge building discourse.

Acknowledgements

The research reported here was supported by the Centre for Research and Development in Learning, Nanyang Technological University (CRADLE@NTU). We would also like to thank participating teachers and students in this study.

References


Using CSCL to Conceptualize Disability Toward Inclusive Education Design

Christopher P. OSTROWSKI
Werklund School of Education, University of Calgary, Canada
chris.ostrowski@ucalgary.ca

Abstract: This paper presents a framework and design activity to support pre-service PSEs (PSEs) in conceptualizing disability and inclusive education using representations within a computer-supported collaborative learning (CSCL) environment. PSEs often only receive limited or cursory training about disability and how to design for inclusive education. Discourses focus on diagnoses and prescriptive support strategies, which can limit PSEs’ ability to effectively support students’ needs. I argue PSEs need opportunities to engage with their epistemological and ontological commitments to disability before advancing to considerations for design. Moreover, how PSEs identify in relation to disability can impact their agency in supporting students’ needs and how they shape social structures of how disability and inclusion is manifested in their contexts. Framed by Dewey’s notion of an aesthetic experience, I propose a design activity where PSEs develop and exercise metarepresentational competencies to individually and collaboratively create, critique, and interpret external representations, such as drawings, imagery, or models of disability and inclusive education. Such representations can be mapped to a digital network and used as mediating artifacts toward collaborative conceptualization of disability. Furthermore, PSEs can draw upon the network to refine, manipulate, and create new representations of possibilities of inclusive education. The network affords the ability for PSEs to democratize their learning and situate their representations, identities, agency, and social structures to see how they collide, converge, and diverge. As noted by CSCL scholars, meaning making does not occur inside the mind and requires external social negotiation of ideas. Combining representational design with CSCL offers a powerful way for PSEs to create socioculturally informed conceptualizations of disability and consider implications for inclusive education design.

Keywords: Inclusive education, representations, disability, teacher education, CSCL

1. Introduction

When an educator is told a student has a disability, what assumptions, approaches, and reactions are triggered? The dominant discourses about disability (Gilham & Tompkins, 2016) focus on diagnoses and prescriptive strategies to manage disabilities. What is missed is an explicit discussion of pre-service educators’ (PSEs) epistemological and ontological commitments to disability and how they manifest in designing for inclusive education. Problems arise when PSEs try to design inclusively without engaging in deeper dialogue of what disability might be, what it might not be, and what it could be. I argue opportunities are needed for PSEs to articulate, represent, negotiate democratize their conceptualizations of disability and how they can affect inclusive education design.

I propose a theoretical framework and design activity to support PSEs in better understanding disability by collaboratively inventing, interrogating, and manipulating representations. I discuss how a metarepresentation design activity (diSessa, 2004) within a computer-supported collaborative learning (CSCL) environment may help PSEs develop more holistic conceptualizations of disability and inclusive education. This representation design activity is framed as an aesthetic experience which is characterized by the transformation of a material into an expressive form, continuity across personal lives and disciplinary perspectives, and an in-depth perception of other’s representations (Dewey, 1934/1980). The framework and design activity are situated within the context of a Canadian Bachelor of Education degree program but could be adapted to other contexts, disciplines, and topics.
2. Theoretical Framework

The proposed design activity aims to surface PSEs’ assumptions, understandings, and perceptions of disabilities through external representations, and to leverage such representations as entry points into discourses about disabilities and their impact on teaching and learning. I draw on several theoretical perspectives, including inclusive education, metarepresentational competencies (diSessa, 2004); and identity, agency, and social structures (Holland, Lachicotte, Skinner, & Cain, 1998; Shanahan, 2009); CSCL (Stahl, 2006); and aesthetic experience (Dewey, 1934/1980).

2.1 Inclusive Education

Teacher education programs traditionally approach disability from a medical perspective (Gilham & Tompkins, 2016), which addresses support funding and accommodations but not inclusive design. Though well-intentioned, psychologists—by design—focus on diagnoses and individual therapeutic treatments. They evaluate students’ strengths and limitations to recommend support strategies based on diagnostic codes (Alberta Education, 2016). Such clinical approaches are a necessary part of supporting student needs but are only part of the solution. Efforts are being made to understanding disability more holistically but are still emergent (Gilham & Tompkins, 2016). There is a touted and accepted vision of inclusion, but a comprehensive training and implementation plan is lacking (Alberta Teachers’ Association, 2014). PSEs typically have positive attitudes toward inclusion but much of the training offered is information-based and focuses on policies and procedures without deeper engagement with the topic and PSEs’ assumptions (Collins, 2013; Redmond & Lock, 2017; Tait & Purdie, 2000).

In my research (Ostrowski, 2016) and experiences working with students with disabilities, PSEs’ assumptions about disabilities are a common barrier students face in having their needs met and being included as productive and valued contributors to learning. Similarly, in a local course about diversity, I observed PSEs struggled with conceptualizing disability and how it affects learning. PSEs often focused on cursory aspects of disability, such as diagnoses and support strategies instead of deeper epistemological perspectives, which caused them to struggle with articulating and operationalizing their concept of disability. If PSEs are expected to design inclusive learning environments and experiences, they need opportunities to articulate, negotiate, and refine the notion of disabilities. It is not enough to approach inclusion based on medical labels of disabilities alone.

Inclusion means embracing the distinct ways of knowing, interpreting, meaning making, and ontological and epistemological perspectives of the world, that students with disabilities may have (Lewis, 2017). As Miele (2017) noted from personal experience and a lifetime dedicated to research and accessibility, inclusion is “far more than…adherence to standards” (p. 6). It is “a design consideration which, in order to be meaningful and effective, must come from a user-centered culture holistically integrated with other critical factors of design and usability” (p. 6). As Seelman et al. (2008), Ostrowski (2016), and others have discussed, much of accommodations and assistive technology are localized at “the body level.” That is, they address the specific need of the user to adapt to a fixed environment. Conversely, universal design approaches address accessibility at the system level to be inclusive of a wide range of users. Traditional discussions of disability have adhered to the former in providing support and accommodations for people with disabilities, while leaving the rest of the classroom environment unchanged.

Following a formulaic approach to supporting students may exclude unique ways of learning and knowing. As well, diverse perspectives should be supported with diverse ways to communicate, engage, and express learning. Tools such as language, visuals, and text mediate ontological and epistemological possibilities of how people think and learn. They also dictate human development: “if one changes the tools of thinking available to a child, his mind will have a radically different structure” (Vygotsky, 1978, p. 126). I argue a shift is needed toward inclusion and diverse meaning making as a necessary part of learning. “Inclusion is not primarily a special education, or even an education, issue. It is a fundamental way of seeing and responding to the human difference for the benefit of everyone involved” (Lawarence-Brown & Sapon-Shevin, 2014, p. 4). To reach such perspectives PSEs can benefit from robust conceptualizations of disability and inclusion to effectively support students.
2.2 Metarepresentational Competencies

I suggest PSEs can develop and use metarepresentational competencies (MRCs) to learn about disability and inclusion more deeply. MRCs can challenge PSEs to go beyond sanctioned notions of disability and help them conceive novel approaches to supporting student needs. Metarepresentational competencies are concerned with how students,

- Invent or design new representations
- Critique and compare the adequacy of representations and judge their suitability for various tasks
- Understand the purposes of representations generally and in particular contexts and understand how representations do the work they do for us.
- Explain representations (i.e., the ability to articulate their competence with the preceding items).
- Learn new representations quickly and with minimal instruction. (diSessa, 2004, p. 293)

Conceived in the context of science and math learning, I adapt diSessa’s notion of scientists as designers of representations of scientific phenomena to PSEs as designers of representations of teaching and learning experiences. Perhaps the most common educator design activity is lesson planning, which represents more than a list of topics and procedures for a given day. Lesson plans are artifacts of PSEs’ ontological and epistemological perspectives melded with prescribed curriculum requirements, and influenced by sociocultural pressures of principals, superintendents, budgets, parents, and students. Simultaneously, PSEs must design ways to accommodate, include, and support students with disabilities among the classroom milieu.

To design for such a complex and emergent system, PSEs must develop an anchored (but not static) understanding of disability and inclusion. MRCs can support this process by having PSEs develop robust representations of disability that can be continually, interrogated, articulated, and refined. Creating and using representations is an integral part meaning making and learning (Azevedo, 2000; Enyedy, 2005; Norman, 1993). Representations are tools to think with and tools to express understanding (Norman, 1993). Exercising MRCs can help PSEs expose blind-spots, bright-spots, and murky aspects of disability and consider implications for inclusive design.

2.3 Identity, Agency, and Social Structures

How PSEs approach disability and inclusion is informed by their identities, agency, and the social structures they work in. According to Shanahan (2009),

- Identity is defined by the requirements, norms and expectations imposed on individuals as a result of their membership and position within a social group
- Structure refers to the underlying principles that shape the normative patterns within social groups
- Agency, on the other hand, refers to each individual’s ability to shape the world around them. (p. 45)

These elements impact how PSEs perceive disability and their roles in relation to students with disabilities. For example, if an educator identifies as someone obligated to correct, cure, or fix a student’s disability their approach will differ from an educator that views disability as part of a continuum of diversity. This affects an educator’s agency in their role in offering supports or advocating for student needs. Likewise, the social structures of schools, communities, and government can strongly influence how PSEs approach disability. PSEs can also shape the structures they work within through advocacy or championing inclusive strategies within their schools and communities.

Complementary to identity, agency, and social structures are figured worlds: “socially and culturally constructed realm[s] of interpretation in which particular characters and actors are recognized, significance is assigned to certain acts, and particular outcomes are valued over others” (Holland et al., 1998, p. 52). Figured worlds shape identity development, where identities “trace our participation, especially our agency, in socially produced, culturally constructed activities” (Holland et al., 1998, p. 40). The worlds are socially produced and dependent on peoples’ roles and positions. Figured worlds are always nested, intersecting and overlapping within and among other worlds. Navigating these figured worlds in the physical and observable sense depend on mediational means such as peoples’ goals, the constraints and affordances of tangible and intangible tools, and relations of power and authority (Wertsch, 1998).

How PSEs use representations to discuss and collectively engage the topic of disability can be viewed through the lens of discourses and utterances (Gee, 2011). PSEs’ approaches to disability
depend on how their representations collide and bump against each other. In Bakhtinian terms, each utterance depends on previous utterances and shapes the utterances that follow (Todorov, 1984). The interplay of these utterances influences how people conceptualize and develop viewpoints of the world (Bakhtin, 1981). Interactions occur through utterances which include two inseparable parts: verbal, the linguistics and vocabularies of language; and nonverbal, the contextual implications such as physical objects, individual knowledge of the situation, and shared evaluations of the situation by the participants (Todorov, 1984).

Together, discourses (utterances), identities, and figured worlds affect how PSEs develop and exercise MRCs. The representations that PSEs design will be influenced by their epistemological commitments, which will inform how they interrogate peers’ representations. For example, some may feel pressured to create representations perceived as politically correct or how status quo suggests disability should be represented, which could differ from their actual attitudes and how they manifest inclusive educational practices.

2.4 Collaborative Learning

The metarepresentation activity involves PSEs articulating, discussing, and negotiating the meanings of representations. Articulation and negotiation will be critical in using MRCs to raise students’ awareness of their perceptions and epistemological positioning relative to disability and inclusion. Moreover, meaning making is an external process that depends on a person’s interactions with sociocultural contexts (Vygotsky, 1978). That is, meaning does not innately form inside the mind and is then externalized by actions and artifacts. Instead, “the meaning fundamentally emerges in the external, observable, intersubjective world of other people and physical objects. The external meaning can secondarily be internalized” (Stahl, 2006, p. 338). For this reason, making meaning of a familiar, yet obscure, concept like disability depends on collaboration to internalize external meanings. Which is not to say students are empty vessels needing to be filled. Rather,

The practices of meaning making are acts of discourse or interaction; these acts propose, negotiate, display, and define what are to count as the salient features of the setting, the occasion, the social norms. Neither the context nor the meanings are objectively given in advance but are collaboratively constituted or brought in. Artifacts are not simply instruments for conveying independent meanings but are themselves embodiments of meaning. The process of embodying meaning in artifacts mediates or transforms that meaning. (Stahl, 2006, p. 332) PSEs’ representations are artifacts and mediational tools that embody meaning. The collaborative meanings PSEs form, depend on and, are situated within the communal efforts of the group. They are reflective of localized cultures, priorities, and figured worlds. The meanings transcend individualistic views and “cannot be reduced to the sum of independent individual contributions” (Stahl, 2006, p. 344).

2.5 Aesthetic Experience

The proposed framework advocates for teaching and learning as an aesthetic experience. Dewey (1934/1980) described aesthetic as more than something being beautiful or pleasing to look at. Though as Farris and Sengupta (2016), and Kosso (2002) noted, this too is an important part of knowledge development and disciplinary advancement. For example, many physicists would agree there is beauty and elegance in the theories of famous scientists such as Richard Feynman and Albert Einstein because of what the work represents. Those that study such work know how transformational they were—and still are—in understanding the universe. Their theories represent the deep interconnections and coherence of disciplinary knowledge.

From a Deweian perspective, an aesthetic experience is one that is transformative, continuous, and perceptive. It is the clarification of dissonance and part of both creation and consumption. In Dewey’s Art as Experience (1934/1980), he described it the context of artistic works, but as he noted, an aesthetic experience transcends disciplinary boundaries. Here, I focus on three aspects of aesthetic experience:

- Transformation: transforming a material into an expressive form
- Continuity: the interconnections across personal experience and disciplinary domains
- Perception: a bodily and emotional connection to an experience
These are relevant to teaching and learning because they democratize teaching and learning. They support PSEs in bringing in their identity, agency, and interests into the learning process to make it personally meaningful.

The goal of continuity is to create alignment between PSEs’ lives outside the classroom and disciplinary content. For Dewey (1934/1980), the perceived strength in this continuity dictates the value of such learning experiences. That is, if PSEs cannot see a strong connection between personal life and school, they are less likely to value it. In this context, the goal is to bring relevance and meaning to PSEs about disability and inclusion by encouraging them to draw upon personal experiences in creating and interpreting representations.

Dewey (1934/1980) discussed transformation in terms of an artist transforming paint and canvas into an expressive medium. More than brushing paint on to a canvas, transformation is about taking disjointed or dissonant fragments and molding them into an expressive form. The result represents interconnections and cohesion greater than the sum of its parts, and something that cannot be disassembled into discrete elements. PSEs’ representations reflect the essence of their epistemic perspectives, identities, and agency in how they understand disability and inclusion. The representations showcase what is personally meaningful to PSEs.

Perception, is a bodily experience where there is an inner commotion and stirring of an organism between what is known and what is unknown (Dewey, 1934/1980). Dewey used the example of visitors in an art gallery that might admire or enjoy art pieces but experience no lasting impressions or consequences. “An act of perception proceeds by waves that extend serially throughout the entire organism” (Dewey, 1934/1980, p. 53). The intent of the MRC activity is to go beyond merely labelling and recognizing disabilities, toward a deep engagement and interrogation of ideas to form robust conceptualizations. As Dewey (1934/1980) stated, “For to perceive, a beholder must create his own experience” (p. 54). This aspect relates to how PSEs interpret their peers representations and disciplinary perspectives. To perceive, they need to engage at a depth comparable to the creator of the representation in order to form their own experience of it.

Together, transformation, continuity, and perception form an aesthetic, a cohesion of components that are called into action and presence in learning. “That which distinguishes an experience as [a]esthetic is conversion of resistance and tensions, of excitations that in themselves are temptations to diversion, into a movement toward an inclusive and fulfilling close” (Dewey, 1934/1980, p. 56). Moreover, in a classroom context this precludes the need for an aesthetic experience to be visually appealing or beautiful.

The proposed framework extends previous MRC studies by highlighting the roles of identities and social structures in the process. Much of existing representation research has focused on mathematics and science (e.g., Azevedo, 2000; diSessa & Sherin, 2000; Enyedy, 2005), which inherently seldom consider how identities and social structures matter for representations. However, more recently, Halverson (2013) extended this work to examine digital art and considered the relationship between identity and representational practices.

3. Metarepresentation Design Activity

The design activity includes five parts based on diSessa’s (2004) MRCs and Stahl’s (2006) cycles of design within a CSCL environment: design of individual representations, critique and interpretation of peers’ representations, collaborative design of representations, networked sharing of representations, and refinement of representations and implications for inclusion. In this section I describe each part of the activity based on a scenario of PSEs within a teacher education program. Figure 1 shows a simplified depiction of this process. In practice, this process is less linear and could include overlaps, and back and forth interactions between elements.
To help illustrate the framework, I draw upon three studies including the work of Selling (2016) and Azevedo (2000) on using representations in STEM based contexts (see also diSessa & Sherin, 2000; Enyedy, 2005), and Haleverson’s (2013) study on digital art to highlight aspects of identity and culture in representations.

Selling (2016) explored how seventh and eighth graders created and used representations as problem-solving tools in an algebra class. The students used representations to construct, reason, justify, and discuss claims about quadratic equations. Over five weeks, student math journals and video data were collected to analyze the types of representations used (tables, verbal, algebraic, graphical), evidence of connections between representations, students’ interactions in creating and reasoning with representations, and students’ overall representational practices. Teachers gave students different algebra problems and asked students to create representations, link different types of representations, and use representations to solve the problems.

Azevedo (2000) studied ninth graders’ ability to create representations of landscapes and geographical terrain. Given a model landscape with various features such as a spherical dome, smooth washboard, dip, and an elliptical mound, students were asked to draw representations of this landscape. The three-dimensional model was meant to simulate what could be seen flying over terrain with various geological and topographical features. The study was motivated by the premises that creating and using such representations is a common part of professional practice, and that students seemed to possess substantial constructive resources (prior knowledge) about representing geographic features on paper. The goal was to use to learn about representing landscapes and develop continuities between the representational practices of students and disciplinary approaches. Azevedo (2000) and his colleagues (diSessa & Sherin, 2000) believed in “developing learning activities that engage students’ abilities and interests, and which are couched in appropriate material and social infrastructures…[and saw the] process of invention and reinvention as key for developing deeper understandings of the topics at hand” (Azevedo, 2000, p. 444).

Halverson (2013) researched how representations could bridge digital art making and new literacies. She examined how youth created representations that demonstrated understanding. She described the process as a representational trajectory where youth developed a story about the self and considered how different tools afforded the representation of that story. Like the STEM based studies, representational practices are a process of forming an aesthetic where particular features are highlighted.
and arranged to convey meaning. Halverson observed and interviewed youth in their creation and use of representations across four youth media organizations. Because the representations were about the participants’ autobiographical stories, they were inherently personal and reflective of the participants’ identities. As Halverson (2013) found, there can also be tension in the production of representations while travelling along a representational trajectory. Creating a representation requires an understanding of the relationship between the material to be transformed and the intended idea(s) to be conveyed through the representation. Since there may be a disconnect between a person’s expressive capacity for a given medium, the proposed framework emphasizes the intended and functional meaning of a representation rather than its visual appeal while still remaining aesthetic in nature.

3.1 Design of Individual Representations

PSEs will design individual metarepresentations using available modes and technology, such as digital or paper sketches, models, imagery, or artifacts. In Dewian (1934/1980) terms, PSEs will transform materials into a form that expresses their perspectives and experiences. The corresponding conceptual resources PSEs use to conceive their representations will also be identified, such as experiencing disability, relatives and friends, media, or courses. By first creating individual representations, PSEs can express their identities, agency, and figured worlds of disability, which supports continuity between their personal and academic lives. Individual representations also give PSEs currency for dialogue during collaboration and mitigates suppression of ideas by more vocal contributors. Creating tangible representations fosters reflexive thinking and requires higher order thinking beyond defaulting to generic definitions of disability (Furberg, Klyge, & Ludvigsen, 2013).

For example, in Selling’s (2016) study each student first created an individual representation in their journals and then discussed each others representations with peers in small groups. This gave each student an opportunity to exercise agency in creating an individual and unique representation. Paralleling Dewey (1934/1980), Azevedo’s (2000) participants transformed material—paper and drawing utensils in this case—into a representation that expressed the original model. Specifically, the students were told “the representations had to be good enough to aid a hypothetical hiker interested in navigating that terrain” (p. 450). In Dewian terms, this implied a degree of aesthetic was needed to create a stand-alone representation that communicated the various geographical features in a cohesive way. Students were free to represent the model however they wished, which fostered continuity by encouraging agency and drawing upon prior experiences.

3.2 Critique and Interpretation of Peers’ Representations

In small groups, PSEs will share their representations, articulate their meanings, critique others’ representations, and analyze how representations align and collide. The representations can act as structuring resources (Furberg et al., 2013) to support dialogue and allow reflection on how PSEs’ understanding of disability is situated among peers, instructors, and canonical sources: “representations are also resources for analyzing information, communicating ideas, and coordinating interaction in collaborative problem-solving processes” (Furberg et al., 2013, p. 45). With intention, PSEs can begin to perceive (in the Dewian sense) other’s representations and develop continuity with their personal lives by forming their own experiences of other’s representations. They can consider the similarities and differences of their figured worlds and the possible connections between them.

After Azevedo’s (2000) participants created individual representations of a landscape, they shared and discussed their work. In evaluating each others representations, students identified criteria for preferring one representation over another, including preferring representations other than their own. Azevedo (2000) noted students preferred three-dimensional drawings because as one student stated “It looks more natural…something you can relate to,” to which another student replied, “It’s something you can really identify with” (p. 458). These comments suggest students perceived (or at least started to) the representations by making personal connections to them.

Sharing and discussion of representations can also lead to students modifying their representations, as was the case for students in Selling’s (2016) study. Selling noted students often used their individual representations as a tool to explain their thinking to each other and reason about mathematical problems. They also questioned each other’s lines of thinking to negotiate their
understanding. Similarly, Halverson’s (2013) participants gave oral presentations about their representations and used reflective critique to push their ideas further.

3.3 Collaborative Design of Representations

Prefaced by the interrogation of each other’s representations, PSEs will collaboratively create representations and identify the conceptual resources used. The intent is to foster intersubjectivity, where “learning can be construed as the act of bringing divergent meanings into contact, and instruction as the social and material arrangements that foster such negotiation” (Stahl, Koschmann, & Suthers, 2014, p. 490). This process prompts PSEs to critically consider and socially negotiate the epistemological and ontological commitments of their representations. The collaborative representations are also artifacts of collective meaning making that can surpass what an individual could accomplish. The resulting representations will reflect a collective aesthetic that encompasses the graphical elements the group believes are needed to convey their message.

Although Azevedo’s (2000) participants only created individual representations, they often discussed and negotiated strategies and criteria for what constituted an ideal representation. Based on these discussions and seeing each other’s representations, students refined or redrew their representations. Similarly, the students in Selling’s (2016) study worked together to compare different types of representations and make connections between them. They iteratively went back and forth to come up with the connections and create new representations collaboratively. Selling noted the frequent use of “I” and “we” among the students signaling the interchange between individual agency and corroboration in their approaches to solving problems and creating representations.

3.4 Networked Sharing of Representations

Individual and collaborative representations from an entire class will be uploaded into a digital network application. Similar to a mind-map, this can surface patterns of representations and the associated resources. PSEs will then have access to the representations created by all peers and groups. This has several benefits: first, teacher education programs can use the epistemological landscape of PSEs to improve instructional practices; second, PSEs can situate their conceptualizations within their local context and see how representations collide, converge, and diverge; third, the map can be used as a mediational tool to foster further discourses and engagement with the topic; and fourth, it affords more chances for PSEs to encounter diverse conceptualizations of disability than through casual interactions alone. The epistemic landscape of representations reflects a collective aesthetic at a class-wide level. The map can show localized conventions of what is deemed salient in terms of inclusion and disability. It is also a proxy for the range of identities and cultures that make up the local context. The course instructor can then foster continuity by prompting PSEs’ to consider relationships between their representations and disciplinary perspectives, as the teacher in Azevedo’s (2000) study did in explaining the disciplinary conventions of topographical maps.

3.5 Refinement of Representations and Implications for Inclusion

Using the network of representations, PSEs can draw upon the repository of conceptualizations to refine and manipulate their representations. Among a larger class, such as my local program, the network could include hundreds of representations. Collaboratively, PSEs can leverage the network to create representations of inclusion and implications for design and supporting students’ needs. Here, PSEs can look beyond the aesthetic of their small groups and consider how other PSEs’ representations compare to their own. Without the artist available to explain their representations, PSE’s will be prompted to perceive (Dewey, 1934/1980) other’s representations and use them to refine their representations.

For example, the teacher in Selling’s (2016) study gave opportunities for students to analyze representations created by another class. Students were tasked with trying to figure out how other students solved problems and their thinking behind the representations. Based on seeing other’s representations, the students created “novel but personally meaningful ways to represent an idea. They modified and elaborated on these representations as they used them to make progress on the problem” (Selling, 2016, p. 204). The students also connected different representations to similar problems to see how different methods could be used to solve them. This process required perceptive practices in trying
to understand and make use of other’s representations. In other words, they formed their own experiences of other’s representations.

In summary, the representation activity fosters an aesthetic experience by having PSEs transform materials into an expressive form, develop continuity between personal experiences and disciplinary practices, and shape their future practices by perceiving other’s work (Dewey, 1934/1980). The activity challenges PSEs to consider their identities and roles in supporting students with disabilities and fostering inclusion. Producing representations requires deep engagement and, at times, a sense of internal commotion and uncertainty. As one of Halverson’s (2013) participants stated, producers of representations “struggle with that their pieces are about. They struggle with who they are and how they’re going to show themselves” (p. 157). Creating representations is a way for PSEs to exercise agency in democratizing conceptualizations of disability and shaping the social structures of what disability and inclusion might be, might not be, and could be. The discourses and interactions can highlight the identities and figured worlds PSEs have and form collectively.

4. Conclusion

This paper proposes using MRCs and CSCL to interpret how disability and inclusion are perceived, understood, and conceptualized by PSEs. Existing approaches focus on diagnosing and prescribing supports corresponding to a diagnosis, which can limit how PSEs understand and support students’ needs. The framework extends MRCs as a collaborative process supported by a digital network and contributes to the use of MRCs outside of science and math (i.e., STEM), where it has typically focused. As well, using the CSCL network, conceptual resources can be clustered and the relations between PSEs’ representations can show how PSEs traverse between individual and group representations (Stahl, 2013).

The metarepresentation activity aims to raise PSEs’ awareness of what disability and inclusion means, and “to represent the complicated relationship between the way they see themselves, the way others see them, and the way they fit into the communities to which they belong” (Halverson, 2013, p. 158). The activity is framed as an aesthetic experience that fosters transformation, continuity, and perception to help PSEs “see more” (Dewey 1934/1980) rather than merely recognizing disability and “grabbing things by their most familiar and convenient handles” (Higgins, 2008, p. 13). An aesthetic experience promotes a democratized learning process that is personally meaningful and that transcends disciplinary boundaries. Through developing and using MRCs to engage with the topics of disabilities, inclusion, and PSEs’ and students’ identities, PSEs may better understand the challenges and opportunities for inclusive practices. The activity can give PSEs pause to consider more holistic and thoughtful ways of designing for inclusion and valuing diversity.

References

Alberta Education. (2016). Special education coding criteria. Edmonton, AB.
Measuring Process and Outcome of the Scientific Argumentation in a CSCL Environment

Wenli CHEN* & Chee Kit LOOI*

*National Institute of Education, Nanyang Technological University, Singapore

*wenli.chen@nie.edu.sg

Abstract: This paper describes the conceptualization and operationalization of the scientific argumentation in a CSCL environment. An online platform is designed to support students’ collaborative argumentation with diagram-based representations of argumentation based on Toulmin’s (1958) Argumentation framework. Based on existing analytic frameworks of collaborative argumentation while accommodating the specific demands and characteristics of the target users and the environment involved, a conceptual framework and a group of indicators are derived for operationalizing the measurement constructs of the process and outcome of students’ collaborative argumentation.

Keywords: collaborative argumentation, computer-supported collaborative learning, learning analytics, assessment for learning

1. Introduction

In recent years there is a shift in science education from focusing on exploration and experiment to the construction of argument and explanation (Duschl, Schweingruber, & Shouse, 2007). Argumentation is part of the practice of science for evaluating, refining and establishing new theories (Duschl, 1990). It has been widely recognized as an effective approach for science learning (e.g. Bell & Linn, 2000; Osborne & Patterson, 2011; Zimmerman, 2007; Zohar & David, 2008) as it helps students improve their conceptual understanding (Jiménez-Aleixandre, Bugallo Rodriguez, & Duschl, 2000; Bouyias & Demetriadis, 2012), understand the nature of science, promotes deeper learning of content (Nussbaum, 2008) and enhances knowledge creation (Erduran, Simon, & Osborne, 2004).

Many effective argumentation happens among students (Scheuer, Loll, Pinkwart, & McLaren., 2010) who engage in proposing, critiquing, coordinating evidence with claims to construct arguments and explanations, reflecting, and evaluating each others’ ideas. Educational researchers have developed a good number of pedagogical approaches and tools to support students’ collaborative argumentation (Scheuer, Loll, Pinkwart, & McLaren, 2010). However collaborative argumentation rarely takes place in school science classrooms. One of the critical issues is that teachers and students lack knowledge on what a desired collaborative argumentation is and how to work towards it. In such circumstances, the diversified measurement and assessment is needed to allow teachers and students to have a good understanding on the current status of the collaborative argumentation learning processes and its compatibility with the desired (Jermann & Dillenbourg, 2008), which in turn brings about more effective and efficient collaborative work.

This paper is to conceptualize and operationalize the collaborative argumentation in science classroom. It is part of a research project which attempts to support students collaborative argumentation in science (e.g., Clark & Sampson, 2007; Osborne, Erduran, & Simon, 2004; Zohar & Nemet, 2002) by providing automated assessments on students’ performance through multi-dimensional learning analytics. It is believed that with appropriate conceptualization, operationalization and measurement of collaborative argumentation, productively collaborative argumentation processes can be engendered, and the students can be guided to function effectively and efficiently by monitoring, evaluating, and adapting their learning during collaborative activities (Gress, Fior, Hadwin, &Winne., 2010; Mora, Caballé, & Daradoumis, 2016).
The paper describes the operationalization of the constructs of various aspects of collaborative argumentation including social participation, interaction patterns, argumentative knowledge construction process, cognitive artefacts and reasoning/epistemic patterns. Relating the process data to the outcome data, meaningful processes for productive collaborative learning can be identified. This will provide insights on relevant pedagogical design accordingly.

2. Literature Review

2.1 Argumentation and Learning

Argumentation is a key human skill that is used in a variety of domains and situations (Scheuer, Loll, Pinkwart, & McLaren, 2010). The production of coherent arguments to justify solutions and actions is critical to solving problems (Cho & Jonassen, 2002). The ability to evaluate and reflect on arguments and counter-arguments is an important component of critical and inventive thinking skills that can enable sound decision making and task performance (Quinn, 1997; Nussbaum, 2008). Argumentation is also viewed as a vital type of knowledge construction activity that can lead to knowledge advancement and improvement (Weinberger & Fischer, 2006).

There are argumentation frameworks and models developed which inform the study. One of the most widely used argumentation framework is Toulmin (1958) model which has been used by many researchers (e.g., Cho & Jonassen, 2002; Erduran, Simon, & Osborne, 2004; Jiménez-Aleixandre, Rodriguez, & Duschl, 2000; Kelly, Drunker, & Chen, 1998; Kenyon & Reiser, 2006; Krummheuer, 1995; Osborne, Erduran, & Simon, 2004). The Toulmin (1958) Model is analytical and provides the structure of an argument which consists of six main components: claim, grounds, warrants, qualifiers, backings, and rebuttals. A claim is an assertion, or statement, about a belief or idea. Grounds are statements or reasons that support the claim. Warrants are an elaboration on the reasoning behind why the person believes their claim to be true. A qualifier provides strength and clarification to the grounds and warrant. With a qualifier, the claim is valid only during a specific circumstance. A rebuttal is a particular condition in which the warrant becomes void and the claim is not valid. While a qualifier can provide strength and clarification, the backing provides support to the warrant by stating why the warrant is acceptable.

2.2 Learning Analytics of Collaborative Argumentation

In existing literature on computer-supported collaborative learning (CSCL), a variety of inter-psychological mechanisms embodied in discourses that favor collaborative knowledge construction have been identified (e.g., conceptual controversies: Johnson, 1981; considering others’ perspective: Järvelä & Häkkinen, 2000; formulating ones’ own point of view: Webb, 1991; progressive discourse: Bereiter, 1994; Oshima, Oshima, & Matsuzawa, 2012; Scardamalia & Bereiter, 2003; exploratory dialogue: Mercer, 2000; Mercer & Littleton, 2007). The discovery and exploration of these pivotal processes for knowledge advancement provide the possibility of leveraging on online learning analytics, i.e. the measurement, collection, analysis and reporting of data about the learner and the contexts (Buckingham Shum & Ferguson, 2012).

Through discourse analysis, the presence or absence of the intended knowledge construction processes can be revealed, which can serve as good indicators for tracking and assessing whether the unfolding learning processes are productive or not from the perspective of co-construction of knowledge. There has already been some exploration that goes for this direction in the field of learning analytics research. For instance, Ferguson and Buckingham Shum (2011) have demonstrated that the attributes of exploratory dialogue such as challenges, extensions, evaluations and reasoning can be automatically identified within online discussion. Following this approach, the emergence and evolution of exploratory dialogues which are most desired by educators can be monitored.

Besides analyzing and visualizing the pivotal processes that can positively contribute to collaborative knowledge construction, the epistemic nature of reasoning enacted by students in the process of argumentation is to be evaluated in the project as well. This is a very important aspect for assessment as it can shed light on the type of reasoning that students use when they propose, support,
evaluate and challenge ideas (Clark et al., 2007). Among the analytical frameworks developed for analyzing argumentation, quite a few laid their focus on the analysis of such epistemic moves made by students that can reflect the type of reasoning processes involved (e.g., Walton, 1996; Jiménez-Aleixandre, et al., 2000). These frameworks can guide the design and development of indicators for measuring the application of reasoning skills and strategies employed by the students. Jiménez-Aleixandre, Rodriguez and Duschl (2000), for instance, established nine categories of epistemic operations (namely induction, deduction, causality, definition, appeals to analogy, exemplar, instance, attribute or authority, consistency with other knowledge, experience, commitment to consistency, or metaphysical, and plausibility) for understanding how students elaborate, reinforce, or oppose the arguments of each other. Analyzing the proportions of argumentative dialogue and epistemic moves made in the dialogue can help reveal how students approach argumentation. Comparing the reasoning/epistemic processes enacted by the students and those “desired” ones (enacted by “experts”), the attainments and gaps in students’ work can be identified, and both quantitative and qualitative feedbacks can be provided to help teachers and students to adjust their following actions accordingly.

3. The CSCL Environment

The project developed a prototype of a computer-based collaborative argumentation system to support students’ collaborative argumentation with diagram-based representations of argumentation (Chen, Looi, Xie & Wen, 2015) (see figure 1 for the system interface). On the diagram-based argumentation space, an argument refers to an organized set of argument elements represented by nodes and/or directed links. The specific types of argument elements adopted are in accordance with Toulmin’s Argumentation Pattern (TAP) (1958, 2003). For pragmatic considerations (e.g., understandability of secondary school students) (Scheuer et al., 2010), the original TAP model is simplified. Three argument elements, namely claim, evidence for (support), and evidence against (rebuttal) are identified as the essential components of an ideal argument. These elements are represented by: 1) the type of Node: Claim vs Evidence and/or; 2) the type of directed Link: For vs. Against.

Figure 1. AppleTree System Interface

4. Conceptualization and Operationalization of Collaborative Argumentation

Existing analytic frameworks that are used for examining the quality of argumentation in different formats (e.g. dialogue-based vs graph-based) and with different focuses (e.g. the process vs outcomes of argumentative knowledge construction; the cognitive vs social processes of collaborative argumentation) were examined for the concepulisation of argumentation in this paper. This extensive
search and research helped identify the critical dimensions of collaborative argumentation that deserve measuring and assessing. Then the dimensions and indicators that reflect our view on collaborative argumentation as a critical type of knowledge-construction discourse, a social practice that can lead to knowledge advancement (Clark et al., 2007). As widely acknowledged, the configuration of educational measurement should reflect not only the specific perspectives on learning but also the pedagogical goals and the environment structure characteristics (Clark et al., 2007). Figure 2 show the conceptual framework of the collaborative argumentation measurement. The collaborative argumentation can be measured from two aspects: social process and cognitive outcome. Basing on existing valid and reliable analytic frameworks of collaborative argumentation (e.g., Erduran, Simon, & Osborne, 2004; Toulmin, 2003; Weinberger & Fisher, 2006; Zhang, Hong, Scardamalia, Teo, & Morley, 2011) while accommodating the specific demands and characteristics of the user and the environment involved, a group of indicators were derived for operationalizing the measurement constructs.

4.1 Social Process

The social process assessment is about measuring the process of the collaborative knowledge construction. Engagement and Centralization focus on social participation and interaction process.

4.1.1 Engagement. Engagement refers to the frequency of contribution to the group work. The higher the frequency, the higher the level of engagement is.

4.1.2 Centralization. Centralization is the degree to which the group members equally participate in group interaction. It is measured by the inequality of interactions by different members within the group. The higher the inequality, the lower the centralization is.

4.2 Cognitive Outcome

The cognitive outcome assessment is concerned with the evaluation of the artifacts that a student or a group of students create when asked to articulate and justify claims or explanations. It reveals the advancement of scientific knowledge and the engagement in the epistemic practices in doing science (e.g., coordinating between theory and evidence; taking alternative perspectives into consideration) and the application of reasoning and epistemic strategies. Structural completeness, relevance, scientific sophistication and epistemic complexity are the four constructs identified for measuring the argument as the outcome.

4.2.1 Structural Completeness. Structural completeness is the presence of the essential structural components in an argument generated. The better an argument is, the more components are included. In the present context, an argument contains a claim, one (or more than one) evidence for, and more than one evidence against is regarded as a complete argument. There are 4 levels of measurement for structural completeness: Level 1 - one claim without any evidence; Level 2 - One claim with one “evidence for” OR one “evidence against”; Level 3 - One claim with one “evidence for” AND one
“evidence against”; level 4 - One claim with more than one “evidence for” AND more than one “evidence against”. The structural completeness of the argument diagram is computed by the sum of scores of all the arguments it contains.

4.2.2 Relevance. Relevance concerns whether the evidence provided is related to the topic under argumentation and whether it can support the claim or the evidence that it is directed to. There are 4 levels of measurement for relevance. Level 1 - Irrelevant information/facts; Level 2 - Some relevance but no logic coherence; Level 3 - Relevant and logic but not reflect the key points; Level 4 - Relevant and logic, and reflect the key points.

4.2.3 Scientific Sophistication. Scientific sophistication refers to the extent to which students have moved from an intuitive toward a scientific framework. Scientific sophistication represents the level of success a student has achieved in processing an idea at a certain complexity level. The higher the sophistication, the more scientific the idea that produced is. There are 4 levels of measurement for Scientific Sophistication. Level 1 - Misconception; naive conceptual framework; Level 2 - Misconceptions that have incorporated scientific information but show mixed misconception/scientific frameworks; Level 3 - Basically scientific ideas based on scientific framework, but not precisely scientific; Level 4 - Scientific explanations those are consistent with scientific knowledge.

4.2.4 Epistemic Complexity. Epistemic complexity refers to the extent to which students make effort to produce theoretical explanations and articulations of hidden mechanisms central to the nature of science (i.e., providing and elaborating explanations or justifications) besides providing descriptions of the material world (i.e., providing unelaborated facts). Epistemic complexity represents the level of complexity at which a student/group chooses to approach an issue. Epistemic complexity of an argument element is measured by the cognitive effort taken to processing it as reflected in the content. The greater the cognitive effort, the higher the complexity is. There are 4 levels of measurement for Epistemic Complexity. Level 1 - Unelaborated facts: Description of terms, phenomena, or experiences without elaboration; Level 2 - Elaborated facts: Elaboration of terms, phenomena, or experiences; Level 3 - Unelaborated explanations: Reasons, relationships, or mechanisms mentioned without elaboration; Level 4 - Elaborated explanations: Reasons, relationships, or mechanism elaborated.

5. Conclusion

The system is a knowledge representation tool where the structure of argumentation is explicitly represented to support students’ collaborative argumentation. To better inform the teaching and learning practices, the measurement should reflect the critical aspects of the collaborative argumentation and addressing the need from teachers and students in classroom learning activity. The measurement indicators established reflect the commonly acknowledged view on collaborative argumentation as social processes that can enable knowledge construction and creation. The future work include validating the measurement framework and indicators by empirical studies in classrooms, and automating the human coding through social network analysis, natural language processing and machine learning technologies etc.

References


Case-based Portraits of Contrasting Micro-Interaction Processes During Online Assessment of Collaborative Problem Solving

Johanna PÖYSÄ-TARHONENᵃ, Esther CAREᵇ, Nafisa AWWALᵇ & Päivi HÄKKINENᵃ
ᵃFinnish Institute for Educational Research, FIER, University of Jyväskylä, Finland
ᵇAssessment Research Centre, ARC, University of Melbourne, Australia
*johanna.poysa-tarhonen@jyu.fi

Abstract: This study recognizes the role and the quality of social aspects in collaborative problem solving (CPS) processes and outcomes. The aim of this study, relying on multiple data and phases of analysis, is to explore and visualise, through contrasting case-based portraits, how micro-interaction processes at pair level evolve during CPS assessments in an online environment. The assessment is designed for a student pair in the STEM domain. The results show that in despite students’ similar CPS performance scores, variations in micro-interactions occurred across pairs. It is expected that studying these patterns at pair level may provide new insights into CPS processes and so to support acquiring these skills.

Keywords: case-studies, collaborative problem solving, computer-supported collaborative learning, directed content analysis, peer interaction, social aspects of learning and teaching, qualitative research

1. Background of the Study

Collaborative problem solving (hereafter CPS) combining critical thinking, problem solving, communication and collaboration (Care, Scoular and Griffin, 2016; Griffin and Care, 2015) has received increasing interest as a 21st century skill suitable for formative assessment (e.g. ATC21STM, www.atc21s.org; OECD’s PISA 2015, http://www.oecd.org/pisa/). However, the assessment practices for complex 21st century skills, such as collaboration and CPS have remained relatively vague (Strijbos, 2011; Vista, Care and Awwal, 2017). The tools or approaches used do not always reflect collaborative learning in a way that takes into account the complexity of cognitive, social and motivational factors as they occur over the collaborative process (Kumar, Gress, Hadwin and Winne, 2010). In addition, a general set of indicators with which to assess the quality of collaborative activities, or upon which to compare students’ collaborative learning activities, have been lacking (Strijbos, 2011).

Our theoretical understanding of CPS builds on the Assessment and Teaching of 21st century project and its extensive theoretical framework for technology-enhanced formative assessment of CPS skills (ATC21S; Hesse, Care, Buder, Sassenberg and Griffin, 2015; Scoular, Care and Hesse, 2017). In this study, CPS refers to a shared activity in dyads or in small groups to transform a problem state into a desired goal state (Hesse et al., 2015). CPS competency can be defined as a set of sub-skills, which consists of five strands of individual and group level capacities under the broad skill classes of social and cognitive skills (Hesse et al., 2015). Following Hesse and colleagues (2015), social skills (i.e. participation, perspective taking, social regulation) are about managing participants (including oneself), referring to the “collaborative” part of collaborative problem solving. Cognitive skills (i.e. task regulation, knowledge building) are about managing the task, referring to the “problem solving” part of collaborative problem solving. “Participation” refers to the readiness to share information and externalize thoughts, “perspective taking” means the ability to take into account the others’ perspectives and, “social regulation” points to the awareness of the strengths and the weaknesses of the group members (i.e. group/team awareness, see Fransen, Kirschner and Erkens, 2011). “Task regulation” is defined as planning and monitoring skills for developing strategies for problem solving and shared
problem representation (i.e. “Joint problem space”, see Barron, 2003; Roschelle and Teasley, 1995) whereas “knowledge building” here refers to the ability to learn and build knowledge through group interaction. In high-quality CPS activity, social and cognitive skills are inherent, may overlap, and are implemented in accordance with situational needs (Hesse et al., 2015). The way in which the participants are able to manage these intermingling aspects of CPS is seen as critical to high quality CPS (e.g. Barron, 2003).

Even though collaboration and CPS are regarded as crucial skills for future learning and are already a part of today’s learning environments, much research on CPS skills in online environments has focused on problem solution by the individual. As described earlier, one part of CPS is the task and the other part is the infrastructure within which the participants create and share knowledge, monitor their progress, and detect and repair the breakdowns in communicative acts that may hamper the evolvement of collaboration (Alterman and Harsch, 2017; Roschelle and Teasley, 1995). This paper concentrates on the latter part of CPS. The process of constructing and maintaining shared understanding of the task at hand is at the core for collaborative learning to evolve (Roschelle and Teasley, 1995; Dillenbourg, Lemaignan, Sangin, Nova and Molinari, 2016). But compared with co-located collaboration, in an online setting the creation of shared understanding requires additional effort from the participants (Dillenbourg et al., 2016), since to regulate interaction and achieve agreement is much more complex (Alterman and Harsch, 2017). In this regard, there may be pressure to achieve joint goals with reduced amounts of sharing (Alterman and Harsch, 2017).

This study recognises the substantial role of social aspects in CPS (e.g. Barron, 2003; Wegerif, Fujita, Doney, Perez Linares, Richards and van Rhyn, 2017). In technology-enhanced assessment environments, current technologies are typically inadequate to the task of automatically analysing the content of online communication. In this study, we move beyond individual skill level to interpersonal processes with focus on the content and the episodes of interaction of collaborating dyads as they participate in CPS assessment processes. The specific aim of this paper is to explore and visualise how micro-interaction processes (Davis, Horn, Block, Phillips, Evans, Diamond and Shen, 2015) in dyads evolve during online formative assessment; especially focusing on the group relational aspects of CPS. It is assumed that in spite of similarity in students’ CPS performance scores, variations in micro-interaction processes may occur across pairs. Studying these patterns may provide insights into CPS and enhance understanding about how to better support acquisition of these skills.

2. Methods

2.1 Participants and Tasks

The research participants ($n = 20$) were students enrolled in a master level teacher education program at a Finnish university. Students were randomly assigned to their pairs. In the study, student pairs completed one bundle of assessment tasks in an online assessment environment, developed during the ATC21S project. In the project, an online environment for formative assessment of CPS skills, based on the aforementioned CPS skills (Hesse et al., 2015) was developed at the Assessment Research Centre at the University of Melbourne. Its primary goal was to maximise the developmental progression of individuals’ skills, such as those in CPS. The environment is based on human-to-human approaches to assessing CPS (Care, Griffin, Scoular, Awwal and Zoanetti, 2015) and comprises a set of online interactive and collaborative problem solving tasks in STEM domains. In the tasks, student pairs are given a unique subset of resources required to solve the problem. To fully understand the problem space and to identify all the necessary resources, students need to rely on their partner (Care et al., 2015; Care, Scoular, & Griffin, 2016). The communication takes place via free form chat interface.

The assessment bundle used in this study comprised four tasks, lasting approximately 60-90 minutes. During the tasks, student pairs (Students A and B) were seated in different classrooms to ensure that the only means of communication was the chat interface. In this study, the bundle comprised following tasks (see Care et al., 2015): “Laughing Clowns” and “Olive Oil”, which are content-free tasks, and “Plant growth” and “Small pyramids”, which are content-dependent. Laughing Clowns, which is the focus of this paper, is a symmetric task whereas the other three are designed asymmetric. Symmetric refers to the characteristic that both students within a collaborative pair are presented with
the same stimulus content and actionable artefacts within the online task space; asymmetric referring to the characteristic that each student within a pair is presented with different information and different actionable artefacts.

In laughing Clowns, two participants are presented with a clown machine and 12 balls to be shared with them. The goal for the students is to determine whether their clown machines work in the same way. In order to do this, the two students need to share information and discuss the rules as well as negotiate how many balls they should each use. The students must place the balls into the clown’s mouth while it is moving in order to determine the rule governing the direction the balls will go (Entry: Left, Middle, Right, and Exit = position 1, 2, 3). Each student must then indicate whether or not they believe the two machines work in the same way. Students do not have access to each other’s screen.

2.2 Data Collection

2.2.1 Objective Measures and Process Data: Auto-Scoring, Activity Logs and Screen Recordings

Students’ work in the assessment environment was assessed individually, with the scoring based on their actions which included movement of artefacts, and the occurrence of chat to collaborate. Students’ completion of the assessment tasks generated log file data (i.e. activity log) and the patterns in these data were automatically coded by the scoring engine according to the Rasch model as indicators of CPS elements, producing information on students’ social and cognitive skill levels (Adams, Vista, Scoular, Awwal, Griffin and Care, 2015). The activity log shows the events between two students working on the task. For example, Table 1 displays the task and player IDs (Task 23, Student A or B), the page of the task (here page 1), actions (e.g. the event type), the contents of the chat (text exchanged), and their timestamps.

<table>
<thead>
<tr>
<th>Task</th>
<th>Page</th>
<th>Player</th>
<th>Event type</th>
<th>Contents</th>
<th>Timestamps</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>1</td>
<td>A</td>
<td>Chat</td>
<td>Do you have any clue what we’re supposed to do here?</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Chat</td>
<td>I don’t know I’m thinking :D</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>A</td>
<td>Chat</td>
<td>Yeah, would you throw first?</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Chat</td>
<td>Let’s throw a ball and see what happens</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Action</td>
<td>startDrag:ball1:70:150</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Action</td>
<td>stopDrag:ball1:70:150</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Action</td>
<td>startDrag:ball1:70:150</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Action</td>
<td>stopDrag:ball1:509:135</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Action</td>
<td>dropShuteL:ball1:509:135</td>
<td>06/10/2015 16.45</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>B</td>
<td>Chat</td>
<td>Where did it come out for you</td>
<td>06/10/2015 16.46</td>
</tr>
</tbody>
</table>

In addition, CamStudio™ software (see http://camstudio.org) was used for the recording of all the screen activity during the CPS sessions. In the beginning of the session, students were informed of the recording intention and were given detailed instructions how to start recording their portal session and how to save their data. During the portal session, the students were also requested to check whether the program was running accurately.

2.2.2 Subjective Measures: Cued Retrospective Reporting

To obtain subjective measures on the CPS processes, the process-tracing method as cued retrospective reporting (CRR) (e.g. van Gog, Paas, van Merriënboer and Witte, 2005) was used, as applied in problem-solving tasks. CRR is defined as a verbal reporting procedure, in which participants are invited to verbalise their thought processes during the task performance retrospectively, based on a cue or cues of their performance (van Gog et al., 2005). In this study, through verbal reporting the aim was to make
explicit the CPS process in terms of students’ self-monitoring of why and how they took the actions, especially as a student pair. In this study, the CRR interview was cued with the screen activity data, including the mouse operations and chat discussions recorded during the assessment session (for an example of a screen capture, see Figure 1). The CRR sessions were videotaped. During the interview, the video camera was directed towards the computer screen (showing the screen capture video recorded during the CPS session) in order to capture the exact point of time discussed. These were also marked down in the transcriptions of the interviews.

![Figure 1. A screen capture from the activity data used as a cue in the CRR sessions (Student A’s perspective, Laughing Clowns).](image)

Table 2: Examples of verbal reports of a student pair acquired in cued retrospective reporting (Student A and B, Laughing Clowns) (Note: Translated from Finnish).

<table>
<thead>
<tr>
<th>Student A</th>
<th>Student B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 1 Laughing Clowns</strong></td>
<td><strong>Task 1 Laughing Clowns</strong></td>
</tr>
<tr>
<td>[…] Well hmm, here, in this exercise I realised at once what it is all about. Uh, maybe I was a bit hasty in the sense that I didn’t fully read this task instruction, that we have some balls in common, but I sort of realised how my computer was functioning. Like one can see it from this, I realised at once what the task is. And then I just tried it out a few times how the thing was functioning and reported [gives a laugh] right away. I don’t know whether one could have like [a short laugh] built some cooperation there so that one would first have told the instruction in the way one understood it. But, on the other hand as I knew who my partner was, I know that [the partner] too is really smart and good at these, so... and then, I just (thought) that she’s likely to get it from there quite quickly as well. […]</td>
<td>[…] Well my partner (didn’t realise perhaps) that I had checked it. Then at one point I still tested and then I wrote down where they go, so that if I drop there from L [Left]. And then I tried to test the same ones [to see] if it goes always in the same way. But then at some point when my partner’s messages appeared there so then I tried with those, and then did it at the same time, and then I wrote it down like that [as letters and numbers in chat] so that where they go from those. Because if I just had looked how they go (--) if I drop from L [Left] for instance, so that it will go to number one, so I would not necessarily have remembered it then. And then I wrote it there already, (---), like it will probably go. […]</td>
</tr>
</tbody>
</table>

2.3 Data Analysis

To better understand how CPS processes unfold at pair level as micro-interactions (Davis et al., 2015), an analysis technique triangulation was used (Humble, 2009; Meadows and Morse, 2001). The analysis relied on the CPS performance measures, obtained from the assessment environment, combined with directed content analysis on the process tracing data (Pöysä-Tarhonen, Care, Awwal, Häkkinen and Ahonen, 2016). Grounded on these results, for this paper two qualitative case-based portraits were chosen and were visualized using activity logs from the assessment environment, notated with the coded CRR data.
2.3.1 Autoscoring of CPS Skill Levels

The assessment environment automatically codes activity logs for individual student performance measures as social and cognitive skills (skill levels between 1-6, see Adams et al., 2015). These data were used for acquiring a general overview of the CPS skill levels of the participating students. At this point, four pairs (eight students) were chosen for further analysis, based on equal technical quality and availability of all the data (including activity logs from the assessment environment, screen recordings, CRR data).

2.3.2 Directed Content Analysis

CRR interviews resulted in qualitative accounts as retrospective reports concerning the social and cognitive aspects of CPS processes from the perspective of an individual student (see Table 2). Directed content analysis that includes the application of conceptual categories to a new context (Hsieh and Shannon, 2005; Humble, 2009) was applied to the transcribed CRR data. The categorization matrix applied the same behavioural indicators of CPS elements as defined by Hesse and others (2015), used in the automated coding procedures in the assessment environment (Adams et al., 2015). In the analysis, the unit of analysis was an episode or passage; a minimum unit, where a certain criterion of the pre-determined category of a particular CPS element was observed. In Table 3 the categorization matrix of Laughing Clowns is presented (see Care et al., 2015 for more examples).

Accordingly, in the assessment environment, the combination of the CPS elements and their behavioural indicators (representing 19 sub-elements) are task-specific, based on the different characteristics of the tasks (Care et al., 2015). Also, not all the elements are present in the different tasks. As a categorizing matrix, the assessment criteria thus included “Social” and “Cognitive” as the main categories and the task specific sub-elements of social and cognitive as subcategories. In addition, if the content was related to CPS but could not be connected to any of the CPS elements in the predetermined coding category, it was placed in a residual category named “Unclassified”. Those parts of the transcribed CRR data where no criteria could be found were left uncoded. In this paper, the residual category was not included in the process visualizations.

Table 3: The categorization matrix of the behavioural indicators of CPS elements in Laughing Clowns.

<table>
<thead>
<tr>
<th>Skill/Element</th>
<th>Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social:</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>Interacting with partner</td>
</tr>
<tr>
<td>Audience awareness</td>
<td>Adapts contributions to increase understanding of the others</td>
</tr>
<tr>
<td>Responsibility initiative</td>
<td>Takes responsibility for progress for the group</td>
</tr>
<tr>
<td>Cognitive:</td>
<td></td>
</tr>
<tr>
<td>Resource management</td>
<td>Manages resources</td>
</tr>
<tr>
<td>Systematicity</td>
<td>Implements possible solutions to a problem</td>
</tr>
<tr>
<td>Relationship</td>
<td>Identifies connections and patterns between elements of knowledge</td>
</tr>
<tr>
<td>Solution</td>
<td>Correct answer</td>
</tr>
</tbody>
</table>

To ensure trustworthiness of the deductive approach applied in this study, a double coding procedure was used (Schreier, 2012). As the code definitions are clear, subcategories do not overlap (Adams et al., 2015). Two rounds of coding by the first author produced, as expected, approximately the same results, which is seen to indicate a good quality of the deductive categorization matrix (Schreier, 2012). Also, to ensure sound interpretation of the data, the coding of the first task of two pairs were verified by a co-author. For data analysis, Atlas.ti®-data analysis software was first used. Next, the coded data were exported as xml to Microsoft Excel for organizing, analysing and visualizing the data. At this phase, based on the categorized data, cumulative frequency distributions were first calculated to summarize the appearance of coded CPS elements by individual students across four different CPS tasks that comprise the assessment bundle. Next, relative frequencies were calculated in pairs across different tasks.
2.3.3 Process Visualizations of the Micro-Interaction Episodes

Even though the ATC21S assessment environment captures all the text exchanged between students via the chat interface, to assess for collaboration over the problem solving processes, the focus has been on placement and occurrence of chat actions during the CPS process (Adams et al., 2015). However, in this study, we broadened our analysis to search for the evidence of collaboration students provide by combining the activity logs with students’ interpretations (CRR data) of CPS processes for tracking the micro-interaction processes (Davis et al., 2015) in pairs. In this study, following Davis and colleagues (2015), the dyadic interactions were not treated as a single interaction thread but seen as accumulation of many periods of interactions of various lengths of time (Davis et al., 2015). In the fine-grained analysis of the contents of CPS processes in pairs, the grounding assumptions were based on the design principles of the particular processes to be activated in a certain task (Care et al., 2015). The aim was to determine the beginnings and endings of these micro-interactions within a student pair. Finally, the visualizations of the micro-interactions were notated with coded CRR data.

3. Results

3.1 How Do Collaborative Problem Solving Elements Appear in Different Tasks (Individual level)?

CPS performance measures of all research participants, based on automated scoring of the assessment environment, did not show strong differences between individual students, especially in social aspects (social skills between were 5-6, cognitive skills between 3-6), therefore not allowing for clearly distinguishing between “successful” or “less successful” pairs in CPS. Directed content analysis of CRR data resulted in frequency distribution of CPS elements of individuals across tasks. The sums of social and cognitive elements of the individual students are presented in Figure 2.

![Figure 2](image)

Figure 2. The sum of frequency distributions of different CPS elements of individual participants in all the tasks. The participants are displayed by their student IDs, such as Student 24a, 24b etc.

Based on the directed content analysis, for this paper two cases (Pair 24 and 29) were chosen for visualizations as contrasting micro-interactions. The CPS skill levels showed only slight difference in skill levels, Pair 29 with lower social skill level (Student A: social 5, cognitive 4; Student B: social 5, cognitive 3) and Pair 24 with higher social skill levels (Student A: social 6, cognitive 5; Student B: social 6, cognitive 3). However, the content analysis of the task Laughing Clowns revealed a lack of social aspects in interpretations of Pair 29 when completing the task whereas for Pair 24 social and cognitive aspects were balanced (see Table 4 for relative frequencies of social and cognitive aspects in pairs). Due to the limited space in this paper, these brief examples are chosen to illustrate the contrast in the most condensed and continuous form, as opposed to the micro-interactions in the other three tasks that would require more comprehensive extrapolation. The examples are, however, representative of the general quality of the data corpus produced by these two pairs across the four different tasks.
Table 4: Relative frequency distributions of social and cognitive aspects, based on CRR interviews (Pair 24 and 29); seen in relation to the designed task elements of Laughing Clowns.

<table>
<thead>
<tr>
<th>CPS elements (Task design)</th>
<th>Social aspects %</th>
<th>Cognitive aspects %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laughing Clowns</td>
<td>43</td>
<td>57</td>
</tr>
</tbody>
</table>

Student interpretations:

<table>
<thead>
<tr>
<th>Pair 24 (altogether)</th>
<th>Social aspects %</th>
<th>Cognitive aspects %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Student B</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pair 29 (altogether)</th>
<th>Social aspects %</th>
<th>Cognitive aspects %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Student B</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

3.2 Micro-interactions in Pairs: Two Contrasting Cases as Joint Solution Endeavour Versus Individual Solution Endeavour

The micro-interactions confirmed these fundamentally different CPS processes (for excerpts, see Figure 3 and 4). The pattern of Pair 24 (“Joint solution endeavour”) depicts an ideal dual-problem space (Barron, 2003; see also Alterman and Harsch, 2017) where participants simultaneously focus and develop the content space (cognitive aspects) and the relational space (collaborative aspects). In the content space they jointly make sense of the significant elements of the task and jointly reason out the task logic whereas in relational space they show being able to manage their interpersonal relations as they collaborate (Alterman and Harsch, 2017). Pair 29, in turn, showed significant challenges in this respect (“Individual solution endeavour”), which leads to self-focused problem-solving trajectories in collaborative context (Schneider and Pea, 2014).

![Figure 3. Joint solution endeavour (Pair 24, Laughing Clowns). The full episode and an excerpt (grey area), including a time-stamped, condensed activity pattern (A= action, C= chat); dialogue extracted from the activity log, speech bubbles contain excerpts from the coded CRR data. (Note: Translated from Finnish)](image-url)
In regard to social aspects of CPS, in Laughing Clowns the fundamental requirement for successful completion of the task is interaction between participants (Care et al., 2015). Participants need to be aware that their balls are shared and that the most effective way of finding the solution is to assign the balls so that both students may have adequate and equal opportunity to trial their machine and reach a conclusion (see Pair 24). However, in contrast to Pair 24, after the opening messages, Pair 29 does not discuss the allocation, but starts working independently, dropping and dragging the balls without interacting and coordinating their efforts by any means, with Student A first. Also, evidence of audience awareness (i.e. how a participant adapts contributions to increase understanding of the others) as another social skill indicator of the task is missing with Pair 29. Students who are proficient in this area are likely to interact with their partner between ball drops and adapt their behaviour to best suit their partner’s needs (in contrast, see Pair 24).

In terms of cognitive aspects of CPS, students with low proficiency of resource management skills may only concern themselves with checking how their machine functions, thereby monopolising use of resources, which seems to be the situation with Pair 29, while more skilled students are likely to recognize the need for shared use of the balls and share them equally (Pair 24) (Care et al., 2015). Moreover, to reach solution, the students need to identify the relationship between entry and exit point of balls and determine if there is a consistency in how the machines function; the pair needs to construct a way of representing this information that communicates to each partner as well as being able to understand other forms of representation that the partner uses. For example, Pair 29, Student A provides a narrative while Student B lists pieces of information. Skilful students will also challenge the patterns and test the assumptions that underpin their observations, which does not occur with Pair 29. The final step comprises the students comparing their representations such that a decision concerning the similarity of clown machine functioning can be made (Solution). Pair 24 communicated throughout their activities, whereas Pair 29 only discussed their shared understanding towards the end. If the task was not forced for collaboration with the last concluding question, Student 29A would have been able to solve the task quite independently.
4. Discussion and Conclusions

The aim of this study was to explore the micro-level interactions in pairs that occur during the online assessment and thereby provide case-based portraits, focusing especially on the group relational aspects of CPS that may otherwise remain unexplored. In small group collaboration, there are multiple interacting elements that contribute to the structure and flow of collaboration, such as aforementioned elements related to joint problem space (Alterman and Harsch, 2017; Roschelle and Teasley, 1995). Significant challenges in collaboration may arise if participants do not attempt to coordinate their individual perspectives into the joint problem space, as witnessed in this study. Participants who sense co-presence perceive each other and pay attention to each other, which is, however, more demanding when students are operating in online settings (Alterman and Harsch, 2017).

Accordingly, the degree of shared understanding the pairs need to reach is related to the task they need to perform (Dillenbourg et al., 2016), which sets requirements for pedagogical design in this regard. But, as Dillenbourg et al. (2016) point out, in collaborative learning, in general, the tasks are primarily designed to facilitate shared meaning and collaboration. However, these practices should not be considered as something that would automatically happen: pedagogical designs can be perceived and interpreted rather differently by different participants (e.g. Arvaja and Pöysä-Tarhonen, 2013). In our example, the symmetry of the task might have diluted the effect of interdependence designed in the tasks and thereby highlights the motivational aspects in regard. In this sense, to better understand how participants produce learning in collaborative situations requires not only to focus on how they build shared meaning, but how they engage in this activity (Dillenbourg et al., 2016). Engagement may appear at various levels, i.e. at behavioural level (i.e. effort and contributing to the task), social level (i.e. the quality socio-emotional interaction and equitable participation) and at cognitive levels (i.e. planning, monitoring and evaluation) (e.g. Sinha, Kempler Rogat, Adams-Wiggins and Hmelo-Silver, 2015). In relatively limited collaborative situations, like the case here, a rather shallow interactional quality might be enough to complete the task. However, in the other tasks studied, being richer and asymmetrical, the other participant’s resources and understanding becomes more critical. Yet, similar patterns of interactional difficulties or even breakdowns can be found with the same dyad; but, due to the increasing complexity, not in such linear modes as presented here.

As our exploratory case study was conducted in a highly structured assessment environment online, characterized by a challenging communication channel (chat) as dyads as the unit of analysis, to be able to generalize the contrasting cases beyond this special research design requires replication with more subjects. In addition, it should be noted that is unclear whether and how the assessment situation, even being voluntary, impacted the ways in which students collaborated. Also, to include novel methods, such as dual eye tracking during working on online collaborative tasks, analysed in respect to identified key interactional events or, focusing e.g. on “micro-monitoring” of partner’s behaviour in dyads (Schneider and Pea, 2013), could enrich the study. But, our case-based approach, with multiple data and phases of analysis that carefully displays processes of dyads building shared understanding and acquiring CPS practices is still rare even in the area of collaborative learning (Schwarz and Baker, 2017; Stahl, 2017). Therefore, through analysing temporal interaction processes in dyads, e.g. by identifying special problems in the appropriation and use of assessment practices online, this approach has the full potential to inform pedagogical design and refinement of practices in this respect.

Acknowledgements

This work is funded by Academy of Finland (Grant no 273970).

References


DBCollab: Automated Feedback for Face-to-Face Group Database Design

Vanessa ECHEVERRIAab*, Roberto MARTINEZ-MALDONADOa, Katherine CHILUIZAb & Simon BUCKINGHAM SHUMa

a Connected Intelligence Centre, University of Technology Sydney, Australia
b ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Ecuador

*vanessa.i.echeverriabarzola@student.uts.edu.au

Abstract: Developing effective teamwork and collaboration skills is regarded as a key graduate attribute for employability. As a result, higher education institutions are striving to help students foster these skills through authentic learning scenarios. Although face-to-face (f2f) group tasks are common in most classrooms, it is challenging to collect evidence about the group processes. As a result, to date, it is difficult to assess group tasks in ways other than through teachers’ direct observations and students’ self-reports, or by measuring the quality of their final product. However, there are other critical aspects of group-work that students need to receive feedback on, for example, interaction dynamics or the collaboration processes. This paper explores the potential of using interactive surfaces and sensors to track key indicators of group-work, to provide automated feedback about epistemic and social aspects. We conducted a pilot study in an authentic classroom, in the context of database design. The contributions of this paper are: 1) the operationalisation of the DBCollab tool as a means for supporting group database design and collecting multimodal traces of the activity using interactive surfaces and sensors; and 2) empirical evidence that points at the potential of presenting these traces to group members in order to provoke immediate and post-hoc productive reflection about their activity.

Keywords: automated feedback; collaboration analytics; interactive surfaces; multimodal; f2f

1. Introduction

Collaborating effectively and knowing how to be a team player have been acknowledged as critical 21st century skills that professionals are required to develop for the future workforce (Bellanca, 2011). Thus, education providers are strongly encouraged to support the improvement of these professional skills. Although there are a wide variety of uses of the terms teamwork and collaboration (presenting quite differentiated characteristics in terms of leadership, roles, shared goals etc. Campbell (2011)), in this paper we focus on those situations where small groups of people need to interact face-to-face to establish common ground in order to complete a specific common task (Dillenbourg, 1999). Indeed, face-to-face (f2f), collocated collaboration activities are not uncommon in regular classrooms and teachers often try to design tasks to nurture skills that can help students learn how to perform effectively in both teamwork and collaborative situations (Guiller et al., 2008). However, teachers often find it difficult to assess and provide effective feedback on group work (Strijbos, 2011). This is partly because it is hard to collect evidence about what happened during the collaborative activity which can serve as a basis to formulate feedback or to support reflection. As a result, teachers commonly provide feedback in regards of the final outputs of the collaborative activity (e.g. final marks, comments, incorrect responses) or based on the limited observations they can make whilst the activity unfolded. This sometimes discourages teachers from considering group tasks as a core component of the assessment and students from being fond of group-based tasks (Pfaf & Huddleston, 2003).

Most of the efforts to automatically capture group’s interactions, with the purpose of providing support and/or feedback, have been developed for distributed or online contexts (e.g. forums, chats, blogs, social networks) (Jeong & Hmelo-Silver, 2010). Some exceptional works have explored the use of interactive surfaces and sensors to capture the complexity of collocated collaboration (see review in Martinez-Maldonado et al., 2016). Authors of these systems have suggested that providing teachers and students with automatically captured evidence about their group dynamics may predict group performance (e.g. Olguin et al., 2009) and support reflection and regulation (e.g. Bachour et al., 2010).
However, these studies have been limited to investigate the provision of feedback only on social aspects of collaboration, lacking information on epistemic aspects.

This paper explores the potential of interactive surfaces and sensing technologies to both support collocated collaboration and provide automated feedback. Feedback in the form of automated assessment and visualisations is provided just after group members complete their task. The aim of this is to provoke group reflection on epistemic and social dimensions of collaboration. We investigate the potential of presenting these analytics of group activity to group members through a pilot study conducted under authentic classroom conditions in the context of collaborative design of databases. For this, we deployed the DBCollab tool, a multi-display environment based on multiple tablet devices connected to an interactive tabletop (see Figure 1, left) that facilitates the collaborative design of Entity-Relationship diagrams. The system also features multimodal sensors (i.e. kinect depth sensor camera and microphone array) that capture group and individual traces of activity. The system provides an interactive dashboard that contains a set of visualisations automatically generated which are presented to learners just after completing their task (see Figure 1, right). In short, the contributions of this paper are: 1) the operationalisation of the DBCollab tool as a means for supporting group database design and for collecting traces of the collaborative activity using interactive surfaces and multimodal sensors; and 2) empirical evidence that points at the potential of presenting these traces to group members in order to provoke immediate and delayed reflection about their activity.

The rest of the paper is structured as follows. Section 2 presents theoretical foundations of feedback and discusses current work focused on the provision of feedback to collocated groups. Section 3 describes the learning context of this work and the implementation of DBCollab. Section 4 presents the pilot study conducted under authentic classroom conditions. Section 5 presents results and discusses the potential of our toolset to generate the analytics means for provoking reflection through automated feedback. The paper finalises with conclusions and a brief discussion of future work in Section 6.

![Figure 1. The DBCollab tool. Left: three learners interacting at the multi-display, multi-touch environment that facilitates the collaborative design of database diagrams. Right: a set of visualisations automatically presented to groups just after completing their task.](image)

### 2. Background

#### 2.1 Foundations of Feedback for Supporting Collocated Groups

Feedback can be broadly defined as any type of information provided by teachers, peers or external agents intended to improve students’ performance (Boud & Molloy, 2013). High quality feedback should include information about: intended goals, current student’s performance, and guidance for developing strategies to close the gap between these two (Sadler, 1989). Feedback has been demonstrated to have positive effects on learning when it is given effectively (i.e. avoiding criticism or negative and personal comments, understandable, specific and selective, balanced) (Nicol, 2010) and timely (immediate or delayed) (Hattie & Timperley, 2007). Reflecting on the feedback provides students with the opportunity to make adjustments for subsequent performance (Boud et al., 2013).

Effective provision of feedback should answer the following questions: what is going to be informed? to whom the feedback is intended? When and why it should be presented, and how this information should be provided? (Hattie et al., 2007). In terms of group work, feedback can be related
to epistemic aspects (e.g. grade, correct/incorrect answers, comments) or social aspects (e.g. participation, contribution, social interactions) (what) (Carvalho & Goodyear, 2014) and can be delivered through verbal, written, visual, audio or video information (how) (Boud et al., 2013). In terms of temporality (when), groups and/or individuals (to whom) expect to receive feedback from their peers, the teacher or any other agent/system during their task (real-time feedback), just after completing the activity (immediate feedback) or in the subsequent hours/days after the collaboration experience (post-hoc feedback). Finally, groups are expected to reflect upon strategies and performance for the next group-work activity (why) (London & Sessa, 2006).

Our work is grounded on these foundational principles to explore the provision of feedback for reflection about epistemic aspects (the task process, goal, outcomes), and social aspects (interaction process, and individual and group outcomes). Visual analytics about social and epistemic aspects of the group activity are automatically generated to provide feedback to group members immediately after the group have completed their work, and for post-hoc reflection days after the activity finalised.

2.2 Providing Feedback to Collocated Groups

Most automated feedback systems for both individual and collaborative learning have been based on computer-mediated systems where all the group interactions are mediated and easily recorded by the support system (Pardo et al., 2017). Whilst this is suitable for non-collocated collaborative situations, these solutions do not provide support for those situations where face-to-face interaction is important. A critical challenge for providing automated feedback in collocated group-work environments is the capture of evidence for generating the feedback. Several technologies (i.e. array microphones, depth cameras, tabletops, wearables sensors) have been used to capture interaction data from collocated environments to analyse activity and mirror information back to collaborators. For example, research work conducted by DiMicco et al. (2007) and Bachour et al. (2010) was based on automatically capturing sound using microphone arrays to reveal speaker participation patterns in face-to-face discussions. In these cases, visualisations, such as bar charts, timelines, bouncing balls, were displayed during the discussion aimed at provoking changes on behaviour. Other work has utilised depth cameras to analyse social dynamics in a group brainstorming activity by estimating the attention of participants using a head tracking algorithm (Schiavo et al., 2014). Overt and subtle directives were displayed during the group activity to each participant to promote balanced participation. This and the previous examples have explored the impact of showing minimalistic feedback in real-time to group members with varied results. These include reports on the possible negative effects of making explicit traces of activity to underperformers resulting in undermining their participations even further.

Substantial work exploiting traces from tabletops has been presented in past years. Evans et al. (2016) automatically modelled social regulation processes occurring in a tabletop during user-centred design activities. Although, this work demonstrated that pattern analysis from touches interaction can reveal the quality of collaboration processes, this information was not presented in any form as a student facing interface. By contrast, the work presented by Martinez-Maldonado et al. (2011) explored the analysis of tabletop traces by integrating a microphone array and a depth camera to discover patterns from the multimodal data. Authors explored the impact of generating visualisations in real-time and showing these to the teacher to orchestrate multiple groups collaborating simultaneously (Martinez-Maldonado et al., 2015).

Our work builds on previous research by making visible some traces of groups activity to students themselves, which has not been deeply explored in previous work. Using similar technology to the one developed by Martinez-Maldonado et al. (2011) to capture traces of verbal and touch activity, our study attempts to explore the feasibility and potential of providing automated feedback on social and epistemic aspects of the group-work activity. In addition, contrary to the studies exposed above, where feedback information was presented during (in real-time) the activity, our study explores the provision of feedback when groups just has finished (immediately after) the activity and, after a while (post-hoc) the activity was performed, with the purpose of aiding reflective discussion and writing.
3. Design and Apparatus

We developed DBCollab, a tool for supporting collaborative database design and reflection about social and epistemic aspects of collaboration through the provision of automated feedback. The next subsections describe the design and implementation of our toolset.

3.1 Design Features of the Learning Environment

The design of the DBCollab tool is based on previous work that focused on defining a set of design features for a multi-touch tabletop system to support argumentation (Falcones et al., 2016) and database design (Wong-Villacres et al., 2015). As a result, we designed DBCollab to support the following features: a) Structuring the task in sub-tasks; b) providing clue-based instructions; c) providing a shared view and highlighting individual contributions; and d) providing a collaborative puzzle-like interface (for more details see Falcones et al., 2016).

3.2 The DBCollab tool and the Sensing Technology

The DBCollab tool was implemented using a) interactive surfaces (a tabletop and tablet devices) and b) sensors (a depth camera and a microphone array).

Interactive Surfaces. Shared interactive tabletop: A 60-inch Ideum tabletop (see Figure 2, IS1) was used to support simultaneous users’ input by touching on the elements displayed at the interface. All actions were automatically logged. Individual tablet devices: Each student is provided with a tablet device (see Figure 2, IS2). Through this personal device, they can add new cards by sending them to the shared view at the tabletop; and ask for, read, and share clues.

Sensors. Identification of the speaker: The microphone array built in a Microsoft Kinect sensor – V.2 is used to identify and record what student is speaking around the tabletop (Figure 1 - S2). The speech and estimated speaking time by each student are automatically recorded. Differentiation of users’ touch: As the tabletop does not differentiate the user touching the interface, a Microsoft Kinect sensor - V.1 tracks student’s hands regarding their position around the tabletop, identifying who touches the table (Martinez-Maldonado et al., 2011).

3.3 User Interface Implementation

DBCollab allows the collaborative co-creation of database designs (diagrammatic entity-relationship schema) for up to three students. The software of the DBCollab tool is composed of three applications: a) the tablet interface; b) the tabletop interface (Figure 3); and c) the dashboard application (Figure 4). The tablet application lets individual members to create, edit cards and send them to the tabletop application through a server. The tabletop application allows students to move, link cards and identify
group member’s actions. All elements created by a student are shown with a different colour (yellow, blue and red). In addition, the tabletop application is connected with a teacher’s application to obtain configuration information and to setup the task. Figure 3 depicts some of the features explained in the previous section. For stage I we can see a snapshot of the tablet device used by group member 1 (red) (see Figure 3, left). She is highlighting words that she considers important from the short description of the case study. During stage III she can ask for a ‘clue’ from three existing clues for this case study (see green button). She can also share the selected clue with group members 2 and 3 (blue and yellow) by tapping on the grey button **Share clue**. Figure 3 (right) shows the tabletop user interface for a group working in stage III. Group members can create and link cards, corresponding to entities and attributes in their data schema. Each card is identified with the group member’s colour.

![Figure 3](image)

**Figure 3.** The DBCollab design tools. Left: the tablet interfaces used in stage I. Right: the tabletop interface mainly used in stage III.

After the design activity, the dashboard application generates both, the epistemic visual analytics elements (see A, B, C, D, E in Figure 4) and the social visual analytics elements (see F, G, H, I elements from Figure 4) of collaboration. Next, we describe each of these feedback elements.

**A. Teacher’s Solution:** This element shows information obtained from the ideal solution proposed by the teacher, with the purpose of informing groups about the expected goal to be achieved.

**B. Group’s Solution:** This element shows the outcome of the group. In this way, students can compare both, the teacher’s and group’s solutions. This information is aimed at encouraging dialogue between teacher and group members to discuss discrepancies.

**C. Replay:** This feature of the interface allows students to replay the partial design solutions from the beginning to the end of the activity. This feature is aimed at supporting awareness about the epistemic process (e.g. showing how the group approached the task and build their final design, step by step). With this information, group members can reflect on how they were going and if the used strategy was useful to achieve the expected goal.

**D. Automatically Generated Grade:** This information was calculated from the degree of similarity between the teacher’s (A) and the group’s solution (B). Providing this information to the group can potentially help them to explicitly evaluate their outcome according to teacher’s expected goal.

**E. Correct and incorrect Entities and Relationships:** Also, by comparing both solutions (A) and (B), we presented correct and incorrect entities and relationships in a detailed list. With this information, group members can reflect on possible misconceptions and mistakes.

**F. Entities/Relationships and group touch’s actions:** (Related to F1, F2 and F3) This information was obtained from touch inputs by counting each time a group member added database elements i.e. attributes and relationships (F1-left, F2) and, counting each time a group member performed a touch action over an element i.e. create, delete and edit (F1-right, F3). Showing this information to the group could provoke reflection about participation at an individual level in the context of other group member’s actions.

**G. Overall Touch and Speech Participation:** Overall participation was obtained from speech and touch inputs, mapped onto a timeline (accumulated participations vs. seconds). Ideally, showing this information to groups can provoke reflection about periods of high and low interaction. Also, this information can be used to help group members reflect on group interactions and how they should improve participation for a further activity.
From Dillenbourg’s (1999) perspective about collaborative learning and the activity-centred analytic framework (Carvalho et al., 2014), Table 1 summarises the type of feedback offered by our dashboard regarding the epistemic and social aspects of collaboration. The expected goal in terms of social aspects of collaboration was not made explicit in the dashboard but it was implicitly stated by the teacher that all group members should equally contribute to the collaborative activity.

Table 1: Type of feedback related to the aspects of collaboration.

<table>
<thead>
<tr>
<th>Aspects of Collaboration</th>
<th>Epistemic</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>A</td>
<td>Implicit</td>
</tr>
<tr>
<td>Process</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Outcome</td>
<td>B, D, E</td>
<td>F1, F2, F3, H</td>
</tr>
</tbody>
</table>

Figure 4. The DBCollab feedback tool. Information generated automatically from student’s interactions, grouped into epistemic and social aspects.
4. Pilot Study

Fifteen undergraduate participants, ranging from 21 to 26 years old, enrolled in an introductory Database Systems subject, were asked to use the DBCollab tool during their regular classes. Participants were organised in triads. The teacher, who was part of the research team (first author of this paper), set two sessions (Sessions 1 and 2) to use the DBCollab tool with students and for the research team to gather evidence about their reactions after receiving automated feedback. Students were asked to sign an informed consent form explaining that all the collected data would be used for research purposes and that no personal information will be shared or used for formal assessment. The duration of each classroom session was two and a half hours. One DBCollab system was used; thus, groups were allocated 30-minutes time slots to perform the design activity. The purpose of the session 1 was to let students to familiarise themselves with the tool, so no automated feedback was provided in Session 1. In Session 2, automatically generated feedback was provided to the groups. The focus of this paper is Session 2, in which groups performed the following:

1. **Collaborative design activity (20 min):** each group was asked to solve the same database design problem using DBCollab. At the end, they presented a final design of the database to the teacher.
2. **Feedback (5 min):** each group was asked to navigate through the dashboard and explore all generated feedback.
3. **Short interview (3-5 min):** a researcher asked students, questions related to the automated feedback. Some of the questions were: Why do you think you got that grade for the final design? Do you think this feedback could help you to reflect on the group-work? Do you think this feedback is useful to help you to reflect on the activity performed?
4. **Questionnaire about perceptions of the feedback:** Each participant filled out a Likert-scale questionnaire with six questions: Q1) potential of feedback for both aspects; Q2) usefulness of feedback for both aspects; Q3) usefulness of feedback about social aspects; Q4) usefulness of feedback about epistemic aspects; Q5) usefulness of epistemic process feedback; and Q6) validity of the automated grade.
5. **Writing reflection (post-hoc activity):** Three days after the classroom collaborative design activity, each student received all feedback results by email and each group member was asked to write a reflective text about the collaborative activity, first as an individual reflection and then a shared group reflection.

5. Results and Discussion

This section presents the analysis of student’s perceptions and reflections about DBCollab. As an initial exploration, we wanted to understand the overall perceptions of students to the tool (questionnaire responses). Then, we present a qualitative analysis of the reflective comments externalised by students during the short interview and the post-hoc reflective writing task.

5.1 Questionnaire Responses: Perception, Validity and Potential of the Feedback Provided

Figure 5 shows the overview of students’ responses to the questionnaire. Our dashboard was seen by most students (71.43%) as an effective support tool to provoke reflection (Q1). Regarding the usefulness of the dashboard for reflection on social aspects of collaboration (Figure 4, F, G elements), 64.28% of students completely agreed with this question (Q2). This may be explained by the fact that no explicit intended goal on social aspects was shown in the dashboard (see Table 1). As students were used to compare their current performance with something more explicit, such as a mark or score, they may have expected to see an indicator of what a good performance in collaboration was. Moreover, 50% of students found our dashboard useful when showing epistemic aspects (Q3). Again, this overall perception from students could be diminished by the lack of assessment about the social aspects of collaboration.

In addition, 64.28% of the students completed agreed that feedback about epistemic aspects of the collaborative activity (see Figure 4, A, B, C, D, E elements) was useful for reflection (Q4). Nonetheless, a small group of students (7.14%), disagreed with this. When analysing comments about
the tool, students stated that the grade needs to be improved, which lead us to the result of the last question (Q6). When exploring the perception of usefulness about the task replay (Figure 4, element C) in the dashboard, only half of the students perceived this feature as useful for provoking reflection (Q5). Students seemed to be more engaged with information presented as summaries (Boud et al., 2013). Finally, regarding the validity of the automated grade and correct and incorrect responses (Figure 4, element D, E), students had lower credibility for the automated generation of this information, for which only 14.29% agreed. This perception could negatively impact the usefulness of epistemic aspects (Q4). Students expressed that the grade should be improved according to the teacher’s solution. More flexibility and less ambiguity should be considered when analysing the similarities of both solutions. This result is related to a recent concern about “algorithmic accountability” (Diakopoulos, 2015), indicating that we need to ensure that algorithms we use should be accountable rather than appearing as black boxes to students.

5.2 Qualitative Analysis of Student Reflections

Although the questionnaire responses provide an overview of the perception of students, we explored more deeply the student’s thoughts and reflections from the feedback provided. First, by analysing the reactions of students when they interacted with the dashboard and the short interviews we gathered some evidence that points at the potential of supporting student reflection by providing timely, automatically generated indicators of epistemic and social aspects of the activity. For this analysis, we tagged the time that every group spent on reflecting about a specific epistemic or social dashboard element. Figure 6 (A) depicts the average time groups spent (min) for both, social and epistemic aspects. For the epistemic feedback, all groups dedicated a significant amount of time to compare their solution (element A) with the teacher’s solution (element B), with the purpose of analysing possible misconceptions about learning performance. All groups, at some point tried to explain why the teacher’s solution had a database design element and why they forgot to add it to their final solution. For example, in the interviews, some of the students expressed this as follows: “What is the meaning of this [pointing to a part of the teacher’s solution]? Oh, I see, that is part of the solution we missed”. Moreover, most students did not reflect deeper on the grade (element D). Some students indicated that the level of detail of the explanation provided in the dashboard is still not sufficient to understand. One student expressed this as follows: “I think there is more detail here [pointing into both, group and teacher’s solution - A and B elements] rather than here [correct and incorrect entities and relationships- E element] ”. In addition, not all the groups (3 out of 5) spent time to explore the replay option (element C). This could explain its low-rating of usefulness in the questionnaire (Q5).

Regarding the social aspects of the feedback, although groups were positive about the importance of showing this type of feedback, they were less concerned with reflecting about it in responses to question Q2. An explanation of this behaviour could be that the epistemic feedback is

![Figure 5. Results from student’s questionnaire.](image)

![Figure 6. (A) Time spent and (B) comments from reflections about feedback.](image)
presented on the first screen, and to explore the social feedback, students had to remember to click a button and then the second screen is presented. Nonetheless, students suggested that the “touch’s actions” helped them to compare their individual actions with those of other group members. This was described by one student as follows: “Look at this! You [pointing to a specific portion of the touch’s actions - element F3] took almost the half part of the pie chart! That is because you don’t let the others participate”. Also, groups made interesting reflections based on the overall touch and speech participation (element G) by identifying moments where they all were engaged with the activity. For example, one student said “My highest participation point was in the first two minutes. Here [pointing to a part of the timeline visualisation] is where the group had more participation”. Some students could also identify when they started to act as leaders. One of these students described this as follows: “… over the time I started to give directions to my peers and then I verified if everything was ok”. A student referred to the timeline to emphasise when another peer was not engaged with the task. For example, one student said: “[name of one peer], where is he [pointing to the timeline]? Look, he is gone”. Finally, some students mentioned that the detailed participation per group member (element F) helped them to compare one with another peer’s contribution to the solution.

Second, we analysed the written reflection (the post-hoc activity) by counting comments about epistemic and social aspects of the performed activity. Figure 6 (B) presents the average number of comments for both aspects. For this reflection activity, the numbers of social and epistemic written comments were much more balanced compared to Figure 6 (A). Regarding the epistemic aspects of the feedback, students commented on their performance with respect to the teacher’s solution. A couple of students expressed this as follows: “we got closer to the teacher’s solution” and “our performance was bad, we had a lot of errors that were revised from the teacher’s solution”. Students also proposed strategies to improve their performance by means of knowledge preparation. For example, two students stated the following: “I should have reviewed the concepts before coming to the class” and “next time, we should come more prepared to classes so we do not spend too much time thinking”.

Social feedback elements helped students to reflect on their participation as individuals and as a group. Some students explained this as follows: “my contributions helped the group to identify most of the elements for the final solution” and “overall, the group’s contribution was balanced which helped us in reaching a good solution”. Students also identified strategies for further improvement of group performance, such as their “need for better organisation” and a more “equal distribution of specific tasks to solve the problems and give voice to all participants”. In short, the reflective writing activity may have offered more time to recall and reflect more deeply about things that happened, not only in the epistemic domain, but also in the social domain (unlike what was observed immediately after feedback was provided, where students focused on the epistemic aspects of the task).

6. Conclusions and Future Work

This work presented our DBCollab tool aimed at capturing collaborative activities from interactive surfaces and sensors. Our tool can generate automated feedback from multimodal data traces, which is presented to groups just after the activity and for post-hoc analysis. We validated the provision of social and epistemic feedback for provoking reflection through the exploration of the tool in a partially authentic classroom scenario. When students explored the feedback immediately after the activity, they seemed more inclined to reflect on their task performance in comparison with the intended performance (e.g. what the teacher expected from them). By contrast, when students were provided with more time to reflect on the group and task performance (post-hoc reflection), a more egalitarian reflection was noticed. They suggested strategies for improving both social and epistemic aspects of the task. This paper should be seen as a first effort in a series of studies to realise the vision of supporting face-to-face collaborative activities by generating student-facing multimodal analytics interfaces that provoke immediate and post-hoc productive reflection.

Future work is directed towards two key strategies. Firstly, in this study, the formal grading was focused on the quality of the data schema design. Future work will better align the assessment criteria with the analytics, so that it is clear to students how the feedback in the dashboards connects to their primary concern (higher grades). Secondly, we will prototype enhancements to the visualisations designed to make the feedback more intelligible, and actionable, by highlighting key areas for attention.
References


Examining Student Learning of Engineering Estimation from METTLE

Aditi KOTHIYAL & Sahana MURTHY
IDP in Educational Technology, Indian Institute of Technology Bombay, India
*aditi.kothiyal@gmail.com

Abstract: Engineering estimation is the determination of any physical quantity up to a specified level of accuracy. It is an important activity done before and during engineering design. Therefore, engineers need to be able to make estimates, but research suggests that even graduating students are unable to make good estimates. This is because they are not trained in estimation as part of their curriculum. We designed a technology-enhanced learning environment (TELE) based on progressively higher order modelling to train students in engineering estimation. We evaluated the TELE to explore what students learn from it. We found that students learned the various reasoning processes involved in performing estimation, recognized the role of evaluation in estimation and the need for practical considerations in estimation. These results have implications for redesign of our TELE to improve student learning of estimation.

Keywords: engineering estimation, technology-enhanced learning environments, modelling

1. Introduction

Consider this problem: “You are participating in an electric car race in which you are required to design an electric car of weight 7 kg with wheel diameters of 4” that can accelerate at 1 m/s² and traverse a track of 10 m without burning out. Estimate the electrical power needed to achieve this performance and the specifications of the motor you will need.” Engineers often make estimates of, and judgements regarding physical quantities such as this during the design process, in order to establish the feasibility of a design or narrow down the set of design choices. While such estimates are necessary in system design (Linder, 1999; Shakerin, 2006; Dunn-Rankin, 2001) they are hard to make because they involve “mastering the complexity” (Mahajan, 2014) of the system by identifying physical quantities that can be safely neglected. Such estimation is typically used to make a decision that allows one to proceed in the design process when faced with lack of information, resources or strategies.

Research into the differences between practicing engineers and graduating students reveals that there is a marked difference between the performance of the two groups on the quality of estimates for quantities like drag force and energy (Linder, 1999). This indicates that students have very little intuition for these quantities. This is not surprising because as Ferguson observed (Ferguson, 1977) “The real “problem” of engineering education is the implicit acceptance of the notion that high-status analytic courses are superior to those that encourage the student to develop an intuitive “feel” for the incalculable complexity of engineering practice in the real world.” Linder differentiates the characteristics of the learning activities of engineering curricula and rough estimation, and demonstrates that the learning that happens in current engineering curricula do not prepare students for rough estimation activities. This is because these learning activities are primarily well-structured in nature while rough estimation is ill-structured. The ability to solve well-structured problems does not transfer to the ill-structured problems (Jonassen, 1997). Therefore, it is important for instruction to be tailored to developing students’ estimation skill explicitly. A literature survey to identify teaching-learning strategies for engineering estimation showed that while there are some guidelines and activities to teach estimation (Mahajan, 2014; Linder, 1999; Shakerin, 2006; Dunn-Rankin, 2001), there is a lack of evidence-based teaching-learning strategies for engineering estimation.

Research suggests that engineering problems are solved by a combination of model-based reasoning and the use of external representations (Nersessian, 2009; Aurigemma et al, 2013). In particular, the use of external representations helps experts detail out their mental models during the estimation process (Kothiyal et al, 2016). Thus estimation is an instance of model-based reasoning.
In this paper, we present our solution to support the learning of engineering estimation, a technology-enhanced learning environment (TELE) called Modelling-based Estimation Learning Environment (METTLE). In METTLE, students solve estimation problems by following the phases of the MBR estimation process. We describe our design of METTLE and its evaluation with the research goal of exploring what and how students learned about engineering estimation after interacting with METTLE. We conducted a lab study wherein students interacted with METTLE and we interviewed them in order to explore their understanding of what estimation means and how it is done. The results give us insight into how to refine METTLE to better support students' learning of estimation.

2. Related Work

Estimation is a type of ill-structured problem (ISP) (Jonassen, 1997) and so, we draw from literature on teaching-learning of ISP solving in order to design a TELE for estimation. Jonassen proposes an instructional strategy for learning ISP solving, with affordances and scaffolds for each step of PS depending on the cognitive requirements of the step. Several researchers have empirically evaluated the role of various affordances and scaffolds on ISP solving skill. For example, the use of concept mapping in the learning of problem solving has been investigated extensively (Stoyanov & Kommers, 2006; Hwang et al, 2014; Wu & Wang, 2012) in TELES and found to be effective for learning. Similarly, the role and nature of question prompts (Ge & Land, 2004) as a scaffold in the learning of ISP solving in a TELE and from a teacher has been studied and found to improve learning significantly.

As estimation is a type of engineering problem, we also consider research-based teaching-learning strategies for engineering problem solving (Woods et al, 1997; Kalnins et al, 2014; Shekar, 2014; Zheng, 2013). The majority of these strategies are based on using problem-based learning (PBL) and project-based learning (PjBL) in engineering classrooms (De Graaf, 2003; Perrenet, 2000). The salient feature of these pedagogies is a problem or project as a means of organizing student learning of concepts and skills, with emphasis on the learning process, self-directed learning and collaboration (De Graaf, 2003). Among these strategies, there have been variations in which students receive instruction in improving specific problem-solving skills along with, or prior to, attempting the problems (Woods et al, 1997; Pimmel, 2001). This is necessary because PBL/PjBL assumes that students possess or will develop in the process of solving, the necessary problem solving skills, and this may not always be true (Woods et al, 1997). What is needed is to identify target skills and give opportunities to students to practice the skills and receive feedback until they have achieved mastery (Woods et al, 1997). Use of question prompts while students solve engineering problems has been found to improve students' problem solving skills (Zheng, 2013). Another strategy (Wankat & Oreovicz, 2015) focusses on developing students' problem solving method by using a seven step problem solving strategy to design instruction. Students learn problem solving by systematically applying the strategy on different types/levels of problems in order to become aware of their problem solving process.

Several authors have recognized the importance of estimation for engineers and the need for developing the estimation skill explicitly among student engineers (Mahajan, 2014; Linder, 1999, Shakerin, 2006; Dunn-Rankin, 2001). They have presented guidelines for activities that can support the learning of rough estimation. Linder recommends teaching conceptual knowledge of estimation, increasing the number of rough estimation activities done by students and including learning activities that have characteristics similar to those of rough estimation activities such as engineering analysis, sketching, building, explaining and diagnosing, where students have to select relevant information and balance different types of information. Mahajan uses a five-step approach in his course titled “The Art of Approximation in Science and Engineering” with the steps a) describe an estimation tool like divide and conquer b) illustrate its application with an example from a particular domain, c) repeat with examples from different domains d) provide practice in the usage of a tool in practice problems and e) present more practice problems without clues as to which tools to use so that students learn to select which tools to apply. In addition, Dunn-Rankin suggests that whenever possible numerical values be tied to everyday physical objects and activities. This helps students develop an intuition for reasonable values for physical quantities. Shakerin recommends that students be encouraged to practice estimation and be made aware of its importance through short exercises with everyday objects and activities.
Literature on teaching-learning of engineering and general ISP solving offers us research-based guidelines on what types of scaffolds and affordances might support the learning of an ill-structured problem such as engineering estimation; for example, the use of question prompts. However, the specific activities, scaffolds and their sequencing in an effective LE for estimation is not clear from this literature. On the other hand, within literature on engineering estimation, there are several heuristics and strategies that instructors have used, but they have not been substantiated by empirical evidence. In this work, we bridge this gap by designing and empirically evaluating a TELE for engineering estimation. In the next section, we describe the theoretical basis of the design of the TELE.

3. Theoretical Basis of METTLE

3.1 Progressively Higher Order Modelling-Based Estimation

By studying experts solving estimation problems, we found that (Kothiyal et al, 2016) experts begin the estimation process using their preliminary mental models of the given problem and detail them until the model is sufficiently rich enough for the estimation purposes (Figure 1). This first model is based on understanding how the system functions (functional model) and the detailed qualitative and quantitative models are developed by mentally simulating the initial mental model. Further, we found that experts periodically evaluate the utility of their models for estimation and revise them if they do not meet this criteria. Hence it is desirable to guide novices through a similar process where they begin with their own mental models, build, evaluate and revise them to create richer and more useful models by interacting with appropriate affordances and scaffolds in the TELE. For these reasons, we chose progressively higher order modelling as the pedagogical foundation of METTLE. Research has shown this pedagogy to be an effective strategy to improve students’ inquiry and learning (Sun & Looi, 2012; Mulder et al, 2011). We call this MBR learning design as “Progressively Higher Order Modeling with Evaluation and Reflection” (ProHOMER). There are five tasks in ProHOMER namely, functional, qualitative and quantitative modeling, calculate and evaluate. The three modeling tasks each have sub-tasks of create a model, evaluate the model and reflect on the modeling activity. In the calculate phase, students choose and evaluate the reasonableness of values, and calculate the estimate. In the evaluation task, students evaluate whether their estimate is reasonable by two standards. Finally students reflect on the entire estimation process.

In order to develop engineering estimation skill, TELE must provide affordances for and scaffold the creation of all three models, namely, functional, qualitative and quantitative models for estimation (Sun & Looi, 2012; Mulder et al, 2011). Therefore, the TELE must trigger and support students’ mental simulation processes. Literature suggests that students have difficulty in doing mental simulation (Hegarty et al, 2003) and tend to proceed directly to building equations (Wankat & Oreovicz, 2015). Hence we propose the affordance of a fully manipulable simulation of the problem system (Lindgren & Schwartz, 2009). Such simulations have also been used in a number of modeling TELEs (Govaerts et al, 2013; Swaak & De Jong, 2001) to improve students’ modeling abilities.

3.2 Scaffolding Modelling-Based Estimation

In order to learn estimation, students need to actively engage in deliberate practice of the MBR estimation process (Ericsson, 2008). Their interaction with TELE has to be carefully designed such that the TELE scaffolds their doing and learning process. There are several frameworks which define the types of scaffolds needed for complex, ill-structured tasks in TELEs (Basu et al, 2015; Quintana et al, 2004; Kim & Hannafin, 2011). Further, students need scaffolds to learn the MBR estimation process.
and for evaluation and reflection.

Research has shown that external representations such as concept maps (Hwang et al, 2014) knowledge maps (Lee et al, 2005), dual maps (Wu & Wang, 2012), conceptual organizers, process maps, argument maps and causal maps (Quintana et al, 2004; Slotta & Linn, 2009) are effective in order to learn ill-structured problem solving and scientific inquiry. These representations facilitate process management, model building and sense-making. In order for students to manage and learn the three phased MBR estimation process, we propose a visual representation of the process, which simultaneously shows the overall structure of the estimation process and the details of each task and sub-task, in order to give students a “forest and trees” view of the process.

The role of evaluation and reflection on the solving of ISPs such as engineering estimation has been discussed extensively in literature (Jonassen, 1997). In addition, students must learn to think about certain aspects specific to estimation problems such as whether the estimate is reasonable by various standards, which parameters can be safely ignored and how to choose numerical values while doing calculations without compromising on accuracy (Mahajan, 2014). Research has shown that students must be scaffolded in order to articulate and reflect on their inquiry (Quintana et al, 2004) and problem solving (Kim & Hannafin, 2011). Elaboration question prompts have been successfully used in ill-structured problem solving to get students to elaborate and explain their thinking (Ge & Land, 2004). Therefore, we introduce a sub-task in each modelling task wherein students evaluate their models for their utility to give the desired estimate and plan the next modelling tasks, as we had also observed with experts. We use question prompts in order to get students to think about their models and estimated values, the specific aspects of estimation problems and the MBR estimation process.

4. Design of METTLE

In METTLE students solve estimation problems by doing the five tasks of the ProHOMER learning design. The current version of METTLE has the following problem which students solve using the ProHOMER pedagogy: “You are participating in an electric car race in which you are required to design an electric car of weight 7kg with wheel diameters of 4” that can accelerate at 1m/s^2 and traverse a track of 10m without burning out. Estimate the electrical power needed to achieve these specifications.” Students have the option of doing the tasks in any sequence they wish and they may iterate and redo tasks until they are satisfied that they have passed the evaluation check for their numerical value and obtained a good estimate. Once the student selects a task, he/she has to do all the sub-tasks of that task before proceeding to the next task. After solving a problem, they reflect on their estimation process. There are affordances available for creating the models and question prompts for evaluation, planning and reflection. In addition, METTLE has general purpose tools for information, drawing and taking notes, simulating the system, mapping the estimation process and a calculator, which are always available to the student. The key features are described below.

![Figure 2. Estimap depicting the progressively higher order modelling tasks and a modelling sub-task](image)
creating the functional model. The modelling affordance for qualitative modeling is a causal map creator and for quantitative modeling it is a drag-and-drop equation builder. The evaluation and planning sub-tasks have a series of question prompts that students have to answer. For example, the question prompt for “Evaluate the functional model” is “Does the model describe how power is generated and used in this system? Explain.”

3. Calculation task: In this task, the student selects numerical values for parameters in their equation and calculates the power estimate. Students are prompted to think about the “reasonableness” of the numerical values and justify them.

4. Evaluation task: In this task, the student evaluates whether their final estimate is of the right order-of-magnitude and comparable to other known values by answering a series of question prompts such as, “What order of magnitude of power do you expect is needed to run a car? Is the power you determined of the expected order of magnitude? If not, what could be the reason?” The students use the prompts to self-assess their estimate and are not provided any feedback by METTLE currently.

5. Simulator: This consists of a variable manipulation simulation (Fig 3) showing the problem system (a in Fig 3), the parameters affecting power in the system (c in Fig 3) and graphs showing the variation of power with each of these parameters (d in Fig 3). The parameters are presented to the student one-by-one in order to constrain their exploration productively (b of Fig 3).

6. Scratch Pad: In this space students can take notes and draw while they read or use the simulator.

7. Info Center: This space has reference material including documents/webpages/videos for the student to familiarize themselves with the problem system.

8. Reflection activity: In this activity, the student answers a set of question asking them to reflect on their own problem solving process, the tasks which they did and the sequence in which they did them. An example question is, “Why did I need to do all these steps (of estimation)?”

5. Research Method

For this evaluation of METTLE our goals are (1) to explore what students learned about engineering estimation and (2) to understand how the features in METTLE enabled this learning. This exploration is necessary in order to identify the additional features and learning activities necessary in METTLE to improve student learning of estimation. Specifically, this study answers the research question, “What and how do students learn about engineering estimation after interaction with METTLE?”

5.1 Research design and participants

As the purpose of this study is exploration, we collected qualitative data in order to examine the nature of the learning. We performed a lab study and participants were eleven students (one female) from second year undergraduate engineering programs, eight from Mechanical Engineering and one each from Aerospace Engineering, Chemical Engineering and Engineering Physics. They were selected by purposive sampling in order to cover a range of backgrounds - departments and engineering curricula –
in order to increase the likelihood of observing diverse aspects of learning. Further, we had a selection criteria that they had all participated in some non-curricular technical activities such as engineering design competitions. The reason was that we had earlier found that solving estimation problems requires a fluidity with application of engineering concepts (Kothiyal & Murthy, 2015) which develops with technical experience and we did not want lack of this fluidity to be a barrier in the development of the estimation skill. The average age of students was 20 years and they were familiar with the use of computers through other courses and labs in their curriculum. One participants’ data was not used as the audio recording was of poor quality and could not be transcribed.

5.2 Data Sources

We collected multimodal data including (i) individual student semi-structured interviews after their interaction with METTLE (audio-taped and transcribed) (ii) Screen recording of students’ interaction METTLE using the screen capture software CamStudio (http://camstudio.org/) (iii) Student generated artefacts during their interaction with METTLE (iv) Video recording of the student while they worked in METTLE. We answered our research question by analysing student interviews and using the screen recording, student artefacts and video to understand and elaborate on participants’ interview responses.

5.3 Procedure

The overall procedure for the research study consisted of the following steps:
• Initial briefing: We briefed participants about the study and its objectives and their consent was obtained for recording their audio, video and computer screen.
• Pre-test: Students solved an estimation problem on paper, independently and without any researcher guidance. However they were allowed to use the Internet to search for resources/information/concepts that they needed. They were allowed as much time as they needed to solve the question.
• Interaction with METTLE: Students interacted with METTLE and solved one estimation problem. During this interaction they were not allowed to use the Internet; however they were free to ask the researcher any questions regarding how to use METTLE or how to solve the problem in METTLE.
• Individual semi-structured interview: After the interaction, we interviewed students using a stimulated recall protocol wherein their screen capture was played back to them and we asked them to describe what they did at each point in the solving process and reasons for their actions. In addition, we asked them questions about the nature of estimation and the estimation process.

5.4 Data Analysis

We employed thematic analysis (Braun & Clarke, 2006) in order to analyze our data and answer the research question. Thematic analysis is an appropriate method for this research question because our goal is to explore the range of estimation learning experiences existing within our data and the features of METTLE which enabled these experiences. The thematic analysis was done by the first author of this paper, a trained researcher in qualitative methods in Educational Technology. Following the methods of inductive thematic analysis we first transcribed the interviews and familiarized ourselves with the data. At this stage, in order to get a better understanding of student responses and their context, especially when they referred to their actions or created artefacts in METTLE, we studied the relevant screen capture, video or artefacts and added these annotations to the transcripts. Then we generated initial codes across the entire data set and collated related codes into categories and themes. Next we reviewed the themes against the raw data for consistency and generated an analysis map. Finally we refined our themes by examining their details and created clear descriptions of them.

6. Results

We found three themes in students’ learning of estimation, elaborated below: (i) estimation is an MBR process (ii) evaluation is necessary for estimation (iii) estimation requires many practical considerations
1) Estimation is a model-based reasoning process

Almost all students understood that estimation is a MBR process with three modelling phases of functional modelling (understanding the system), qualitative modelling (identifying the relationships between parameters) and quantitative modelling (forming an equation) that must be done stepwise as described here by S4, “You first build up a functional model, that you think of, you only imagine right, imagine the moving parts, like what goes where, what happens when, what pushes what and all sorts of things. You kind of think of an imaginary model, you try to think of an animation, and then you try to get to the various relationships between the quantities, the, qualitative modelling shows exactly that, then the quantitative modelling, there you actually start writing down the equations and tinkering around with them.”

Further students realized that they typically follow this process only sub-consciously, and so tend to skip steps as mentioned by S2, “…the sequence I would say, actually, I always try to follow the sequence in some or the other way, like I generally try to follow the sequence, I might usually miss the qualitative part, but after understanding the requirements, I generally jump to the mathematical part.” Also, students recognized that following the process explicitly would improve accuracy in case of unfamiliar problems, but decrease speed in the case of familiar problems as explained in this quote from S3, “…but if there is something, which I think I don’t know much about that, then, now, I think I should prefer this way, where I would make a model and everything, but, otherwise, if I know the system well, then, I think I will go with that, because, I think that is much, like that would save time for me.”

Students reported that the ProHOMER structure presented in METTLE, via the Estimap, helped them recognize that estimation is an MBR process as described by S1, “So, I went through this [pointing to Estimap] so I knew that evaluation needs to be the last and so …functional modelling was something which I found to be the best part to start with because you need to know how a car runs. Before solving a problem I should know that. After that the qualitative and then the quantitative and the calculation and evaluation.”

2) Evaluation is necessary for estimation

Students recognized that evaluating the models, parameters and values, checking whether obtained estimates are reasonable and realistic and verifying their intuition, assumptions and approximations is necessary in order to obtain better estimates as exemplified in this quote from S5, “I think just that the evaluation part was very critical, because, if that is not there, we might not be able to identify where we have gone wrong at all, and that helps you go through the cycles faster.” Further they also observed that when estimating on their own they usually do not evaluate as S7 mentioned, “Okay, yeah, most of the times when we solve the question, we get an answer, we don’t think that way, that can it be that much or no? But yeah, maybe after solving this, after writing this answer, we would go back and think that, no it can’t be that much and maybe we should do it again.”

Students recognized that the evaluation questions at specific points in the ProHOMER pedagogy, provided by METTLE were critical in getting them to evaluate and subsequently make revisions in their solution if necessary. This is exemplified by this quote from S1, “I wouldn’t have evaluated it at any step or something like that. So that’s where the evaluation part helped me out. It helped me... it stopped me at the crucial places and made me decide like what I have done is right or not.” However, while students learned the need to evaluate periodically and the questions they must ask themselves while evaluating, their responses show that their numerical evaluations were often inaccurate, perhaps because of their limited knowledge and experience. For instance, students often were unable to evaluate whether a particular value was of the right order of magnitude or not.
3) Estimation requires many practical considerations

Students recognized that doing estimation requires thinking of several practical aspects of engineering problems. One of these is quantifying losses, which they have difficulty understanding, as S7 describes: “I didn't even know that there were two different things, we are considering the losses, I didn't know that... that should be made clear that there are losses considered and by what factor is the difference between the input and output power.” Students observed these differences in the simulator, however they were unable to quantify them suitably as seen in their choices of values during calculation.

Students observed that in order to decide which parameters are critical, they need to understand the limits of system performance when it is actually working. As S2 says, “So, it depends on what is critical, in the system where I want to put it, so, if I were to say, the current in the system at certain time, should not exceed the maximum current value, then I cannot go with average value, I need to know that at all times the current is below the Imax, it should not be that current is below the Imax, so, it becomes, a system constraint, that in the system which I want to put it, I cannot have something, like an average is below the constraint, but instantaneous can cross it, so, I think my system will decide how I would use it.” Making such judgements requires experience with similar systems and students often struggle with this as was seen in their choices of values while calculating estimates.

Finally students understood the need to make assumptions and approximations because estimation does not require precision, but speed. However the assumptions and approximations should be reasonable in that they do not cause large errors in the estimate. Students, however, are unable to make these judgements and may end up making inappropriate assumptions for the wrong reasons (such as ignoring air drag when it cannot be), as S5 says, “Umm, because I think that the order of those terms is complex. Like when we were trying to study for them, I think the coefficient of drag that you have to calculate, that depends on a lot of things, and to calculate that I need lots of data, so, I left them out.” The evaluation questions in METTLE alerted them to these considerations, however METTLE did not help them make better judgements as seen from their responses to the evaluation questions.

7. Discussion and Conclusion

The goals of this evaluation of METTLE were to explore what and how students’ learned after interacting with METTLE. In order to answer our research question, we did thematic analysis of the transcripts of student interviews annotated with their created artefacts, and on and off screen actions during problem solving with METTLE. We identified three themes in students learning namely, (i) estimation is a MBR process (ii) evaluation is necessary for estimation and (iii) estimation requires many practical considerations. The analysis also gave insights into how students learned these aspects from interacting with METTLE. Students described the role of the “Estimap” in making the MBR process explicit and this is consistent with the benefits of external representations for process management documented in scientific inquiry (Quintana et al, 2004) and problem solving (Hwang et al, 2014). Also, students were not inclined to apply all the steps of this MBR process for problems that they perceived to be straightforward. Despite the focus questions of the modeling and evaluation sub-tasks, we found that if they were familiar with the problem, students skip the initial steps and jump to equation building, as reported in literature (Wankat & Oreovicz, 2015). This shows that additional scaffolds are necessary in METTLE in order to trigger students’ cognitive mechanism of mental simulation to ensure that they begin with functional modelling regardless of their familiarity with the problem.

Students explained that the periodic evaluation questions in METTLE helped them in learning to periodically evaluate their models and values, and the specific structure of the evaluation question guided them regarding what to consider while evaluating their models. For example the question, “Does your equation relate power to the speed of the car?” highlighted to students that they should think of the speed of the car as one of the parameters. This result is in agreement with research into the
role of question prompts (Ge & Land, 2004) in ISP solving. However, the evaluation questions for numerical values were insufficient for students to learn how to select numerical values. This is because this requires extensive experience and intuition of similar objects and values (Mahajan, 2014). For instance, many students were not able to evaluate what the power required by the electric car would be compared with the power required by a vacuum cleaner because they were unfamiliar with the power of the latter. As recommended in literature, we propose to add activities and additional scaffolds in order to develop students’ skill in choosing and evaluating numerical values, (Linder, 1999; Mahajan, 2014).

Finally, students recognized that estimation requires them to think about many practical aspects, again due to the questions in the evaluation and plan sub-tasks of each modelling phase. But it appears that these questions were insufficient for them to learn how to reason about these practical aspects. This was evident from student generated artefacts in METTLE; students were unable to quantify the losses in the system, justify appropriately why some factors like friction and air drag can be ignored or decide which parameters are critical and which are not. This is because such reasoning is based on experience with similar systems and operating conditions, and the current version of METTLE does not have any guidance on how to think about these aspects nor do students have any exposure to such problems in their engineering curriculum (Mahajan, 2014). We will add scaffolds and activities in METTLE to train students in reasoning about the practical aspects of estimation.

Together these results offer some guidelines to teachers who want to teach estimation in the classroom. Students learn estimation by learning to apply the three-phased MBR process which begins with functional modelling by mental simulation. Mental simulation is a cognitive tool which can enable students to visualize the working of the entire system and how different parameters “flow” inside the system. Teachers can scaffold their mental simulation by providing appropriate question prompts to students which guide them to visualize the layout of the system, think about how it works, what is the mechanism that drives it, what are the aspects that you can control and so on. Further teachers should intermittently prompt students to evaluate their models and values. In order to develop students’ sense of numerical values, teachers must make them do several small activities of comparing values of commonly used physical parameters such as power, force, etc. Finally, doing several such engineering problems will improve students’ ability to reason about the practical aspects of engineering problems.

In future work, we will employ these results regarding what and how students learned in order to understand the cognitive mechanisms that support this student learning. We will do deeper process analyses of student actions during the pretest and interactions with METTLE in order to identify the cognitive mechanisms which enable students to use the external representation of “Estimap” in order to understand the MBR estimation process. Further, process analysis will also show us how students approach ill-structured problems and thus, where and what scaffolds can modify this process in order to become more productive. For instance, we can identify where scaffolds are needed in order to trigger students to do mental simulation rather than equation building.

The sample size of this study is small, which is a limitation. However the larger goal of this evaluation is a rich and in-depth characterization of how students learn about estimation in METTLE, and how this learning mechanism can be made more productive. The current results, which are a part of the larger evaluation, reveal the nature of learning that happened in METTLE. Coupled with the deeper process analyses of student interactions, we will develop an account of students’ cognitive mechanisms as they learn in METTLE, along with what and where additional activities and scaffolds are needed in order to make learning more effective.

References
Social Media Facilitated Group Performance: 
An Investigation of Tie Strength in Grouping

Cong QI
The Hong Kong Polytechnic University, Hong Kong
*cong.qi@polyu.edu.hk

Abstract: The wide adoption of social media has encouraged university teachers to consider employing social media as a new e-learning platform. This research aims to find the role of social media usage in promoting effective communication and enhancing group task performance. What is more, to give implications on a better grouping method, we utilized the concept of tie strength (in a typical social network) in the current research. We proposed that the tie strength among university students negatively moderates the relationship between social media usage and communication. The social constructivism theory was employed as a major theoretical foundation, built on which a research model was developed. The model was tested via 135 senior-year undergraduates in one of the diversified classes. The data analysis results revealed that social media usage promotes communication in group; tie strength in the social network group negatively influences group communication. This means, contrary to our natural assumption, the closer the relationship is among group members, the less likely communication would happen in the group. Our research gave implications on pedagogy in that it extended our understanding of the project-based learning in small teams supported by interactions via social media. It also contributed to the grouping strategies of the classes that rely heavily on the group discussion via social media.

Keywords: Social media, communication, tie strength, group task performance

1. Introduction

The usage of social media is becoming an effective tool in the process of teaching and learning and has increasingly improved the quality of learning outcomes in higher education. Social media are low-cost and low-barrier tools for both students and teachers. From the students’ perspective, a considerable proportion of the university students are “digital natives”, who become frequent users of social media to personalize content and share/participate online (Prensky, 2001). When social media are introduced to the classroom (by digital natives), they become an informal platform to facilitate learning. From the educator’s perspective, social media is contributing towards a disruptive change in pedagogy, known as Pedagogy 2.0 (McLoughlin and Lee, 2011). Pedagogy 2.0 is a framework to achieve learning outcomes by exploiting the potential of Web 2.0 technology. It emphasizes on collaboration, personalization, and user-generated content. Due to the advent of Pedagogy 2.0, instructions in the classroom are moving from the traditional “teacher-centered” approach to the “student-centered” approach (Farkas, 2012). Due to the increasing importance of the pedagogic shift derived from the massive adoption of social media, it is important to explore the consequences of social media in pedagogy.

One of the notable features of social media is their ability to assist communication by real-time information exchange. In the educational context, the increase in communication and interaction is most likely to be attributable to the use of web technology (Andersen, 2004). Social media as a computer-mediated communication platform are expected to promote online connections, maintain relationships and boost communication between students. For example, Ioana (2013) argued that social networking sites like Facebook responds well to the particularities and requirements of the student-centered approach, where students are encouraged to create and develop their own learning style. They further help to promote peer communication, collaboration and active learning among the students. Besides supporting communication, social media as a pedagogic tool can also influence team or task performance in Education.
Tie strength in the social network influences the quality and level of communication (Gilbert and Karahalios, 2009), it is therefore necessary to consider the impact of social tie in the present research. The concept of tie strength was first introduced by Granovetter (1973) when examining the strength of interpersonal ties in the social network. He defined the strength of a tie as “a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie” (Granovetter, 1973, p.1361). Acquaintances and friends with loose relationship are weak ties while trusted friends and family are strong ties. Studies have shown that weak-tie social networks are more effective than strong-tie social networks for sharing information and enhancing social activities (Granovetter, 1973; Levin and Cross, 2004). The primary reason is that people with strong ties (due to their homogeneity) have too many overlaps which reduce the need to communicate; while people with weak ties may have a better chance to acquire and synthesize diverse pieces of information through interactions (Granovetter, 1973). Research has shown that weak ties can help friends to generate creative ideas via communication (Burt, 2004).

Granovetter (1973) originally developed the theory of tie strength, and it has been a useful framework for explaining various social support phenomena such as computer-mediated support groups, virtual community support, and support networks within organizations (Wright and Miller, 2010). Computer-mediated networks are a particularly useful resource for connecting to weak ties who are usually not physically close in a face-to-face relationship (Walther and Boyd, 2002). The emphasis of the weak tie relationship is especially meaningful in the online social networking context, since weak ties could facilitate in receiving different viewpoints, reducing risks, accessing to objective feedbacks from others, and minimizing role obligations (Wright, 2012). From a pedagogic view, studying tie strength within the group is also meaningful, since it could become a new grouping strategy for university teachers to consider before assigning online group works.

To sum up, this paper focuses on addressing two important research questions: does the usage of social media influence communication among group members and their task performance; and does tie strength moderate the relationship between social media usage and communication.

2. Literature Review

2.1 Social Media and Communication

Researchers have extensively studied the effect of social media application on students’ communication and collaboration (Hung and Yuen, 2010; Yaros, 2012). Neier and Zayer (2015) illustrated that students were willing to use social media in education due to the nature of the increased interactivity, which was proven to be a primary motive of social media usage by digital natives (Yaros, 2012). Greenhow and Robelia (2009) demonstrated that the usage of social media could enhance relational support and maintenance and encourage self-presentation. Ellison (2008) also depicted social media as a “social lubricant”, which provides a cost-efficient way for self-presentation and broadcasts personal events to promote interaction and connections (compared with face-to-face interactions). Similarly, Vural (2015) believed that social media, with the ability to contact multiple people instantly, can better notify students of announcements, and facilitate communication, discussion and self-evaluation. Social media are also expected to facilitate communication between students and instructors (Sturgeon and Walker, 2009).

2.2 Social Media and Task Performance

Social media also have a significant effect on task performance. Junco et al. (2011) observed that although two groups of the students had similar high school GPA, the group with Twitter usage in class demonstrated higher engagement and more increase in GPA. In addition, through a study on German students, Skiera et al. (2015) discovered that students located in densely connected subnetworks earn better grades, and this is especially true for male students. Social media can also improve students’ academic performance though enhancing communication, facilitating student engagement and collaboration (Faizi at al., 2013). The positive relationship between social media usage and academic performance has also been proved by the following studies: Isidore (2016) and Lusk (2010).
Nevertheless, there are also opposite views among scholars that believe social media usage is negatively related to the academic performance (Michikyan et al., 2015). For example, Huang (2014) directly mentioned that social media addiction and its symptoms had a significant negative impact on adolescents’ academic performance and social capital. San Miguel (2009) found out that the GPAs of Facebook users were typically lower than those of the non-users. Choney (2010) further illustrated that the usage of social media is a major reason for distraction that impairs students’ academic performance. Using social media for study purposes involves multitasking which undermines students’ capacity to process information and engage in deeper learning (Wood et al. 2012). What is more, Walsh et al., (2013) reported that students who spent the most time using social media had fewer academic behaviors (e.g., completing homework and attending class), a lower academic confidence and more problems affecting their school work. Besides the negative effect of social media, there is also a proportion of the studies that suggest no relationship between students’ social media usage and the academic performance for general purposes (e.g., Junco, 2015; Lambić, 2016). To clarify the ambiguity in the literature, the present study proposed a positive relationship between the two, and empirically tested it in survey.

2.3 Tie Strength

As social media enable students to communicate online in their social networks and social tie is a crucial factor in the social networks, it is necessary to consider the tie strength when investigating the effect of social media usage. In general, people with strong ties are emotionally interdependent and typically provide trust and emotional support to each other. Strong ties were also said to be more effective in tackling complicated projects (Hansen, 1999) and form teams for information dissemination (Shi et al., 2007). Compared with strong ties, weak ties are the individuals that are less emotionally attached. They function as the “bridges” that connect different social circles, support information diffusion and provide access to diverse source of information (Granovetter, 1973). Although strong ties tend to provide more emotional support, the information overlap due to the gravitation towards homogeneity limits the inflow of different viewpoints (Botwin et al., 1997). In contrast, weak ties could enhance the creativity by providing non-redundant information and more complexed information processing behaviors (Perry-Smith, 2014). In the online social network, as mentioned before, weak ties are more likely to provide objective feedback and innovative ideas while the same ability is restricted by strong ties due to the interdependent nature of the relationships (Wright, 2012). Due to the above, weak ties could also facilitate cooperation (Melamed and Simpson, 2016) and knowledge creation (Wang, 2016).

2.4 Grouping Method

In higher education, the group project as an assessment mode is becoming popular, it is necessary to explore the approaches to form better groups and achieve better academic results. Over the years, there has been a sufficient discussion or debate on the right grouping strategy, especially in the setting of the primary school or secondary school. For example, the debate on the ability-based grouping (Magnus, 2016), which classifies the students based on their level of ability or achievement (this includes the sub-categories of Homogeneous and Heterogeneous grouping) (LeTendre et al., 2003); and the age-based (Elizabeth, 2015) or gender-based (Lentillon-Kaestner and Patelli, 2016) grouping, which groups the students based on age or gender differences. For the within-class grouping method (Lou et al., 1996), the literature has suggested grouping students based on ability, gender and friendship (Blatchford et al., 2001). There has been a relatively rare discussion on how the university students should be grouped, especially in the new social environment (the group discussion via social media) and based on the criteria of social ties. As social interaction and relationship building among group members are very important to a successful academic result, the factor of tie strength should be considered before assigning group works to the students. In pedagogy, there has been a limited and/or indirect discussion on the effect of social ties (e.g., centrality and peer effect).

Prior research has investigated the effect of social media and social ties in the educational context. However, there are still gaps in the extant literature: 1. although many studies have examined the effect of social media usage, most of such studies focused on the process aspect of social media usage, e.g.,
engagement, collaboration, and communication (Hung and Yuen, 2010; Faizi, et al., 2013). Few attentions were paid to the study of the social media’s impact on the academic outcomes (e.g., task performance, learning outcomes, grade, etc.) (Mingle and Adams, 2015); 2. there is an ambiguity on the exact nature (positive/negative/no correlation) of the relationship between social media usage and students’ academic performance; 3. there is a scarcity of the studies on the effect of the tie strength when investigating students’ behavior of social media usage; 4. prior research have studied the relationship between face-to-face communication or online communication and students’ performance, few of them have investigated the relationship under the social media context. In view of these, the present research intends to investigate the effect of social media on not only the social process (communication), but also the direct academic performance of the students. We would further introduce the concept of tie strength as a moderator in the relationship between social media usage and communication.

3. Theoretical Foundation-Social Constructivism Theory

Unlike other learning theories that focus solely on how individuals construct knowledge (Piaget, 1953), social constructivism emphasizes the factor of social interaction (Vygotsky, 1978) while learning. In recent years, there has been an emerging trend of constructivism research focusing on the role of social technologies and social media in facilitating the generation of socially constructed knowledge (e.g., Mbati, 2013; Gaytan, 2013). The application of social media in teaching is likely to generate the learning conditions suggested by the social constructivism theory. In other words, social constructivism delineates the importance of social interaction in knowledge construction and social media can help create a social environment depicted by the social constructivism theory. In the present study, social constructivism theory provides necessary theoretical support to the hypotheses development between social media usage and its consequences (communication and task performance).

4. Research Model and Hypotheses

Based on the above discussions of the literature review and the theoretical foundation, we present the research hypotheses and model in Figure 1.

- **H1**: Social media usage is positively related to communication in group
- **H2**: Communication in group is positively related to the perceived task performance
- **H3**: Social media usage is positively related to the perceived task performance
- **H4**: Tie strength negatively moderates the effect of social media usage on communication in group, so that a higher level of tie strength is associated with a lower level of communication, whereas a lower level of tie strength is associated with a higher level of communication.

![Figure 1. Research model](image)

5. Data Collection and Design

The survey instruments were developed based on the literature review and the formal discussions with faculty members who have used social media as teaching and learning tools in their classes. We use 22 items (with 5-point Likert scale) to measure the four constructs in the research model. Social media
usage is defined as the frequency for people to perform social interaction and information exchange on the social networking sites (Hughes, et al., 2012). The students were expected to answer how frequent do they use social media in general to finish the tasks. We used ten items from Lowry, et al. (2006) to measure the communication within the groups (e.g., after using social media, our group communication is appropriate.). Tie strength is used to assess the level of closeness of the relationship between the group members. This research borrowed the (five) questions from Gilber and Karahalios (2009), which was among the first to map the social media data with the concept of tie strength (e.g., I have a strong relationship with most of my group members). Perceived group task performance measures the effectiveness of completing the group project as perceived by the group members. We collected the group members’ self-reported perception of their group project grade in the survey.

Data were collected from one of global strategic management courses in Hong Kong. At the beginning of the semester, the lecturer asked the students to form into a group of five to six people by themselves. Each group was requested to create a Facebook group and invite the subject lecturer to join their groups. Facebook was employed as the official learning platform for students to learn and for teachers to assess their learning outcomes. Through Facebook group, students were required to discuss their projects, respond to group members’ comments and share or upload audio, video, music, or files related to the project (nevertheless, the students could also use other communication tools for informal discussions). They would then deliver a final project report based on the communication via the Facebook platform. The final report would be assessed by the format of a group presentation. The lecturer, on the other hand, served as facilitator and monitored the process by providing feedback, answering questions, and assessing milestones that had been established to ensure groups were on the right track throughout the semester. In the last teaching week of the semester, the predesigned questionnaires were distributed to 150 students in three classes of the subject. After the data cleansing, 135 completed questionnaires were used for the final data analysis.

6. Data Analysis

6.1 Measurement Model

Partial least squares (PLS) structural equation analysis was used to test the measurement and the structural model. SmartPLS (v. 3.2.4), as a mature and widely used PLS software, was employed to test the research model. Table 1. shows that all the composite reliability values are greater than the accepted value (0.70) (Fornell and Larcker, 1981) and the square roots of the Average Variance Extracted (AVE) are more than the recommended threshold (0.5) (Hair et al., 1998). These results demonstrated a good reliability, consistency and convergence of the constructs. Table 2 shows the loadings and cross loadings of the indicators, which confirms the discriminant and convergent validity of the constructs.

Table 1: Composite reliability.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>No. of items</th>
<th>Composite reliability</th>
<th>Square root of AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social media usage</td>
<td>4</td>
<td>0.82</td>
<td>0.73</td>
</tr>
<tr>
<td>Communication</td>
<td>8</td>
<td>0.93</td>
<td>0.79</td>
</tr>
<tr>
<td>Tie strength</td>
<td>5</td>
<td>0.81</td>
<td>0.68</td>
</tr>
<tr>
<td>Task performance</td>
<td>2</td>
<td>0.90</td>
<td>0.90</td>
</tr>
</tbody>
</table>
Table 2: Loading and cross loadings of the indicators.

<table>
<thead>
<tr>
<th></th>
<th>Social media usage</th>
<th>Communication</th>
<th>Tie strength</th>
<th>Task performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMU_1</td>
<td>0.666</td>
<td>0.200</td>
<td>0.006</td>
<td>0.124</td>
</tr>
<tr>
<td>SMU_2</td>
<td>0.896</td>
<td>0.462</td>
<td>-0.095</td>
<td>0.332</td>
</tr>
<tr>
<td>SMU_3</td>
<td>0.696</td>
<td>0.242</td>
<td>-0.068</td>
<td>0.162</td>
</tr>
<tr>
<td>SMU_4</td>
<td>0.631</td>
<td>0.239</td>
<td>-0.080</td>
<td>0.131</td>
</tr>
<tr>
<td>COMM_1</td>
<td>0.374</td>
<td>0.822</td>
<td>0.198</td>
<td>0.389</td>
</tr>
<tr>
<td>COMM_2</td>
<td>0.325</td>
<td>0.875</td>
<td>0.288</td>
<td>0.466</td>
</tr>
<tr>
<td>COMM_3</td>
<td>0.336</td>
<td>0.816</td>
<td>0.118</td>
<td>0.422</td>
</tr>
<tr>
<td>COMM_4</td>
<td>0.288</td>
<td>0.828</td>
<td>0.151</td>
<td>0.347</td>
</tr>
<tr>
<td>COMM_5</td>
<td>0.437</td>
<td>0.848</td>
<td>0.147</td>
<td>0.357</td>
</tr>
<tr>
<td>COMM_6</td>
<td>0.385</td>
<td>0.744</td>
<td>0.184</td>
<td>0.358</td>
</tr>
<tr>
<td>COMM_8</td>
<td>0.193</td>
<td>0.633</td>
<td>0.224</td>
<td>0.325</td>
</tr>
<tr>
<td>COMM_9</td>
<td>0.315</td>
<td>0.682</td>
<td>0.178</td>
<td>0.251</td>
</tr>
<tr>
<td>TS_1</td>
<td>-0.075</td>
<td>0.224</td>
<td>0.751</td>
<td>0.155</td>
</tr>
<tr>
<td>TS_2</td>
<td>-0.191</td>
<td>0.021</td>
<td>0.497</td>
<td>0.162</td>
</tr>
<tr>
<td>TS_3</td>
<td>-0.127</td>
<td>0.173</td>
<td>0.741</td>
<td>0.196</td>
</tr>
<tr>
<td>TS_4</td>
<td>-0.032</td>
<td>0.136</td>
<td>0.675</td>
<td>0.237</td>
</tr>
<tr>
<td>TS_5</td>
<td>0.014</td>
<td>0.140</td>
<td>0.711</td>
<td>0.207</td>
</tr>
<tr>
<td>TP_1</td>
<td>0.333</td>
<td>0.449</td>
<td>0.219</td>
<td>0.922</td>
</tr>
<tr>
<td>TP_2</td>
<td>0.175</td>
<td>0.396</td>
<td>0.270</td>
<td>0.883</td>
</tr>
</tbody>
</table>

Notes: SMU=Social Media Usage; COMM=Communication; TS=Tie Strength; TP=Task Performance

6.2 Structural Model

Table 3 presents the results of path coefficients in the research model. 3 out of 4 hypotheses are significant. The path coefficients of H1 (between social media usage and communication) and H2 (between communication and task performance) are significant at 0.01 level, and H4 (between the interaction effect and communication) are significant at 0.05 level. No significant path was found between social media usage and task performance. As for the R-square (shown in Figure 2), the values for the two important dependent variables in the structural model are 0.28 and 0.23 respectively. This means social media usage and the interaction of social media usage and tie strength contribute to 28% of the variance in communication in group; and all the independent variables together explain 23% of the variance in the perceived task performance.

Table 3: Path coefficients.

<table>
<thead>
<tr>
<th>Paths</th>
<th>Path coefficient</th>
<th>T-statistics</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Social media usage → Communication</td>
<td>0.47</td>
<td>6.06</td>
<td>Yes (0.01 level)</td>
</tr>
<tr>
<td>H2: Communication → Task performance</td>
<td>0.42</td>
<td>4.68</td>
<td>Yes (0.01 level)</td>
</tr>
<tr>
<td>H3: Social media usage → Task performance</td>
<td>0.11</td>
<td>1.30</td>
<td>No</td>
</tr>
<tr>
<td>H4: Interaction effect → Communication</td>
<td>-0.12</td>
<td>2.20</td>
<td>Yes (0.05 level)</td>
</tr>
</tbody>
</table>
7. Discussion

The data analysis results provided support for most of our hypotheses in the research model. First, the results indicate that social media usage is a significant predictor of communication in group. This result is consistent with the prior pedagogic research that social media assist in students’ communication (Hung & Yuen, 2010), increase the interactivity within groups (Yaros, 2012), foster a sense of communication community (Vural, 2015) and enhance relational support and self-presentation (Greenhow & Robelia, 2009). In addition, this relationship was proved to be well-supported by the social constructivism theory which emphasizes the critical role of the social environment in students’ learning process. Second, students’ communication via social media was found to positively influence the perceived group task performance. This means the more communication made among the group members, the better the group will perceive about the outcome of their group work. This result is in line with the literature on communication in general (e.g., Pöysä-Tarhonen, et al., 2016) and communication in the social media context in particular. Third, contrary to our hypothesis, social media does not directly influence students’ perceived task performance as a group. The data analysis results showed no significant relationship between social media usage and perceived group task performance. This result is consistent with a smaller portion of the past research (e.g., Junco, 2015; Lambić, 2016) that there is no significant difference in grades between the frequent social media users and non-frequent social media users for general purposes. What is more, the result is also not upheld by the theory of social constructivism, which promotes a positive role of social media in students’ learning. One possible explanation could be: students may use social media tools for the purposes other than achieving serious academic goals. In other words, the sole use of social media may not affect students’ academic performance, but “how” students use them will. If the students use social media for engaging or information-retentive purposes, the use itself can lead to an enhanced level of perceived task performance (Wang et al., 2012). Last, our results provide some evidence that tie strength among the group members will negatively moderate the effect of social media usage on communication in group. This means social media’s impact on communication in group will be stronger when group members have weak ties with each other. This result is consistent with Granovetter (1973) and Perry-Smith (2014)’s prediction, and demonstrates/reconfirms the power of weak ties in the group-based social networking environment, and in the educational field in particular.

8. Theoretical and Pedagogic Implications

First, it is among the few studies that simultaneously investigated the effect of social media usage on students’ learning process and learning outcome. Second, the general social media usage among students will not necessarily lead to the enhanced task performance. This result indicates that the effectiveness of using social media in education depends largely on the way (e.g. via a serious and high quality of communication) and the objective of using them. Third, prior studies discussed intensively on
the role of communication in promoting online and offline group work; however, few of them examined
the nature of communication among group members in the social media context. Fourth, this research
applied the concept of tie strength to the educational field. It is among the first to emphasize the role of
tie strength, especially the weak ties in the communication process. This research brings a new
perspective to computer-supported collaborative learning by considering the relational closeness of the
students. Last, the social constructivism theory was employed as the theoretical foundation. The social
constructivism theory has a long history in the educational field; however, the application of the theory
under the social media context is relatively new (e.g., Gaytan, 2013). The present study confirms the
explanatory power of social constructivism theory in social media’s effect on communication in group.
For the pedagogic implications, this research extended our understanding of the project-based learning
in small teams that was conducted online and supported by interactions via social media. Specifically, it
contributes to the grouping strategies of the classes that rely heavily on the online group discussions via
social media. When forming project groups, besides the homogeneous or heterogeneous considerations
(Lou, Abrami, Spence, Poulsen, Chambers, & d’Appollonia, 1996), educators should also consider the
relationships between the students.

9. Limitations and Future Research

First, the cross-sectional nature of the survey could only capture a snapshot of the research issues at a
given point in time, but could not depict the evolutionary process of some important constructs (e.g., tie
strength and communication in group) in the research model. We would suggest employing the
longitudinal research design with data collected over multiple periods. Second, the cross-sectional
survey is prone to the common method bias due to the single informants of the survey. Further research
should obtain multiple types or sources of data points to avoid the common method bias. Third, since
the phenomenon somehow involved group-level of the concepts (e.g., communication in group and
perceived task performance), we suggest future research advance the research model at the group level
and conduct multi-level data analysis to examine the relationships among the constructs across levels.
Fourth, though we have tried to diversify the background of the students in the class, it would be ideal to
study the tie strength in the business or broader social environment where respondents may have a real
physical distance, diversified personalities, characteristics, nationalities and relationships, and rely
heavily on the social media to collaborate with each other. The current research design restricts the
scope of studying tie strength into a smaller group of the students who most likely have already known
each other. Last, for the practical consideration, this study only tested the causal relationships with one
type of the social media tools (Facebook), further studies should consider testing the same model with
other famous social media platforms (e.g., google plus) to re-confirm the research results.

10. Conclusions

This study seeks to contribute to the growing body of research by proposing a framework for evaluating
the effectiveness of social media usage among college students and the influence of the tie strength in
students’ group communication. The research results corroborated the findings of the past research that
social media usage significantly influence the communication effectiveness and communication in
group affects the perceived group task performance. A major contribution of this study lies in the
moderating effect of the tie strength. The results provide educators with new insights on the usefulness
of social media in pedagogy, as well as the way of assigning students into different project groups based
on the closeness of the relationships in the social network.

References

practice of online learning (pp. 33-60). Athabasca, Alberta, Canada: Athabasca University.
Blatchford, P., Kutnick, P., Clark, H., MacIntyre, H. and Baines, E. (2001). The nature and use of within class


Impact of Both Prior Knowledge and Acquaintanceship on Collaboration and Performance: A Pair Program Tracing and Debugging Eye-Tracking Experiment

Maureen VILLAMOR\textsuperscript{a,b,*} & Ma. Mercedes RODRIGO\textsuperscript{a}
\textsuperscript{a}Ateneo de Manila University, Philippines
\textsuperscript{b}University of Southeastern Philippines, Davao City, Philippines
*maui@usep.edu.ph

Abstract: We compared the collaboration of pairs whose composition was based on both prior knowledge and degree of acquaintanceship as they traced and debugged fragments of code. We performed a cross-recurrence quantification analysis (CRQA) to build cross-recurrence plots using the eye tracking data and computed for the CRQA metrics, such as recurrence rate (RR), determinism (DET), entropy (ENTR), and laminarity (LAM) using the CRP toolbox for MATLAB. Findings revealed that high prior knowledge pairs who were poorly acquainted (BH/PA) performed better among categories despite having collaborated the least. This confirmed the findings of prior studies that skilled strangers perform best. Mixed prior knowledge pairs who were highly acquainted (M/HA) collaborated the most but their familiarity did not translate to better performance. The results of this study could contribute to the learning sciences and pedagogy. If we know what makes collaboration successful as measured through their performance, we can design interventions that could facilitate the process of creating programming pairs who can collaborate and perform better.

Keywords: Eye-tracking, collaboration, cross-recurrence quantification analysis

1. Introduction

Prior knowledge from previous courses can influence student achievement (Hailikari, Katajavuori and Lindblom-Ylanne, 2008) so it is a known fact that students with high prior knowledge outperform students with low prior knowledge in problem solving tasks. It can also be assumed that when students with high prior knowledge are paired or grouped together in collaborative learning situations, they will perform better than pairs or groups consisting of low prior knowledge students. However, is this always the case if we put friendship into the picture? Previous studies had looked into the impact of friendship on collaborative and competitive performance. They tested whether the quality of social interaction between friends as opposed to non-friends influence collaborative success. Findings have shown that groups composed of friends may perform better as a direct result of their collaboration history (Rittenbruch and McEwan, 2009) or may increase commitment to goals of the group (Jehn and Shah, 1997), which could contribute to more successful collaboration.

However, there is also evidence that friendships could diminish performance because friends have a tendency to focus more on socializing than the group task (Shah and Jehn, 1993). Friendship may even be immaterial for effective collaboration. Groups composed of skilled strangers will perform best because highly skilled individuals may already know from experience how to work well with other experts, and hence, are easily adaptable to the actions of their group mates (Shah and Jehn, 1993).

Pair programming is a collaborative work arrangement where two programmers execute different programming activities together. It has become a well-known pedagogical practice for teaching introductory programming as it has shown that students who are involved in pair programming produce better quality of code, are more confident with their solutions, and are more likely to succeed and persevere in their programming courses compared to solo programmers (Murphy et al., 2010).
In recent years, dual eye-tracking in the context of pair programming has been explored to study joint attention in collaborative learning situations. Two eye-trackers, for instance, can be synchronized for studying the gaze of two individuals collaborating to solve a problem and for understanding how gaze and speech are coupled (Pietinen et al., 2008).

Cross-recurrence quantification analysis (Marwan and Kurths, 2002) is used to quantify how frequently two systems exhibit similar patterns of change in time. It produces a cross-recurrence plot (CRP), which has been used to analyze the coordination of gaze patterns between individuals and can be used to measure how much and when two subjects look at the same spot (Nüssli, 2011).

This paper used CRQA to characterize collaboration of pairs of novice programmers in the act of tracing fragments of code and debugging in a remote pair programming setup. Specifically, it attempted to answer the following research questions: What characterizes collaboration between pairs of participants who (a) both have high prior knowledge that are highly or poorly acquainted, (b) both have low prior knowledge that are highly or poorly acquainted, and (c) have a high and low prior knowledge that are highly or poorly acquainted?

Our prior work focused on characterizing collaboration of pairs based on prior knowledge and acquaintanceship separately using CRQA and seeking for existing patterns. This paper attempts to substantiate the findings from our previous studies by investigating the impact of both prior knowledge and acquaintanceship on the collaboration and performance of the pairs of novice programmers while they traced and debugged codes.

2. Methods

The study was conducted in three private universities in the Philippines. Students aged 18-23 years old who were in their 2nd year to 4th year level in college and had taken the college-level fundamental programming course were recruited to participate in this study. Twenty four (24) pairs of participants were asked to read fragments of code with known bugs and then identify the location of each bug. For a detailed description of the structure of the study, see Villamor and Rodrigo (2017).

To conduct a cross-recurrence analysis, an \( N \times N \) matrix called cross-recurrence plot (CRP) is built, which is essentially a representation of the time coupling between two time series. The horizontal axis represents time for the first collaborator \((C1)\) and the vertical axis represents time for the second collaborator \((C2)\). Recurrence occurs when the distance between the fixations of the two collaborators has to be lower than a given radius. In Figure 1.a, let us assume that the numbered red and green dots are from the fixation sequences of \(C1\) and \(C2\), respectively. Given a certain radius bounded by the black bordered circle shown in the figure, fixation pairs \((1, 10)\) and \((2, 10)\) are considered recurrent since their distances fall within a certain radius.

If fixations \(i\) and \(j\) are recurrent, they are represented as a black point (pixel) in the plot (see Figure 1.b). Hence, a point in the plot indicates that the states of the two systems for their respective times are recurrent. If two collaborators uninterruptedly looked at two different spots on the screen for the entire interaction, the resulting CRP would be completely blank (white space in Figure 1.b). If the two collaborators looked at the same spot on the screen continuously, the plot would show only a dark line on the diagonal. Points exactly on the diagonal of the plot correspond to synchronous recurrence, such as, collaborators look at the same target at exactly the same time.

CRQA defines several measures that can be assessed along the diagonal and vertical dimensions. For the diagonal dimension, we have: recurrence rate (RR), determinism (DET), average \((L)\) and maximal length \((LMAX)\) of diagonal structures, and entropy \((ENTR)\). For the vertical dimension, we have: laminarity \((LAM)\) and trapping time \((TT)\). The definitions of these metrics can be found in Marwan and Kurths (2002).

The number of fixations per slide that contained the actual program were segregated and saved on separate files. A CRP was constructed for each pair for every program using the CRP toolbox for MATLAB (Marwan and Kurths, 2002), and CRQA was performed to get the RR, DET, ENTR, and LAM for each of the 12 programs. The CRQA metrics \(L\), \(LMAX\), and \(TT\) were not included due to the page limit restriction. For this data, no further embedding was done (Iwanski and Bradley, 1998) and the delay was also set equal to one since no points were time delayed (Webber and Zbilut, 2005). The radius was set to 5% of the maximal phase space diameter (Schinkel, Dimigen and Marwan, 2008).
Pearson’s correlation was performed to determine relationships between the categories’ performance task score and CRQA metrics based on both prior knowledge and acquaintanceship. ANOVA was performed twice: (1) comparing the CRQA metric means per program, and (2) comparing the CRQA metric means of the overall task (12 programs). Tukey post hoc tests at 0.05 level of significance were performed to determine which relationships were significant.

![Figure 1](image)

**Figure 1.** (a) An Illustration of Recurrent Fixations and (b) An Example of a Cross-Recurrence Plot.

3. Results and Discussion

A student has high prior knowledge if his/her program comprehension test result was equal to or greater than the median score; otherwise, the student has low prior knowledge. The post-test pair evaluation was used to assess the degree of acquaintanceship of the pairs. Pairs were highly acquainted if their average post-survey rating is within 3.6 to 5; otherwise, they were poorly acquainted.

Of the 23 pairs (one pair was discarded), there were six (6) both high prior knowledge pairs who were highly acquainted (BH/HA), two (2) both high prior knowledge students who were poorly acquainted (BH/PA), eight (8) were mixed prior knowledge pairs who were highly acquainted (M/HA), and three (3) were mixed prior knowledge pairs who were poorly acquainted (M/PA). The remaining four (4) pairs who both had low prior knowledge students were highly acquainted so they were not included in this analysis. The CRQA metrics for every program in these categories were averaged separately to get the aggregated CRQA metrics. Incidences of high and low values of the CRQA metrics were examined to find the differences among the categories. A value is high if it is equal to or greater than the mean plus one standard deviation; and low, otherwise. Table 1 shows the descriptive values of all the aggregated CRQA metrics per program and the ANOVA results per program and overall task.

3.1 Recurrence Rate (RR)

Half of the RR’s in the BH/PA pairs were low and the rest were average. The M/HA pairs had RR’s ranging from average to high. Both the BH/HA and M/PA pairs had one high RR each and the rest were average (see Table 1 for high and low RR). This suggests that the BH/PA pairs collaborated the least while the M/HA pairs collaborated the most. The extent of collaboration of the BH/HA and M/PA pairs in terms of RR were comparable.

One possible explanation for this is because it might be difficult for pairs with both high prior knowledge students who are not familiar with each other to open up for collaboration due to differences in ideas or plainly because secure people are already confident being on their own so they do not really feel the need to collaborate with others. We speculate that this might possibly be the reason for the poor RR result of the BH/PA pairs.

Sharma, Jermman and Nüssli (2012) describe convergent and divergent phases of collaboration. A convergent episode is one in which collaborators look at the same part of the program in a manner that is reflected by fixations less than a given threshold. A divergent episode is one in which collaborators look at the different parts of the program, which happens when participants try to build
their own understanding of the program. This implies that participants are attempting to build their own understanding of the program. We hypothesize that this is probably what happened when the pairs tried to work independently. Their divergent episodes caused their RR’s to drop. However, despite of lower RR turnout, the BH/PA pairs performed the best among categories confirming the findings of prior studies that skilled strangers perform best.

Table 1: Descriptive values and ANOVA results of each CRQA metric

<table>
<thead>
<tr>
<th>CRQA Metric</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Low &lt;=</th>
<th>High &gt;=</th>
<th>ANOVA Per Program</th>
<th>ANOVA Overall Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>0.09</td>
<td>0.03</td>
<td>0.03</td>
<td>0.18</td>
<td>0.06</td>
<td>0.12</td>
<td>9.269</td>
<td>4.700</td>
</tr>
<tr>
<td>DET</td>
<td>0.32</td>
<td>0.08</td>
<td>0.14</td>
<td>0.48</td>
<td>0.24</td>
<td>0.39</td>
<td>14.659</td>
<td>5.500</td>
</tr>
<tr>
<td>ENTR</td>
<td>0.56</td>
<td>0.18</td>
<td>0.17</td>
<td>0.87</td>
<td>0.38</td>
<td>0.74</td>
<td>8.246</td>
<td>4.636</td>
</tr>
<tr>
<td>LAM</td>
<td>0.36</td>
<td>0.12</td>
<td>0.10</td>
<td>0.59</td>
<td>0.25</td>
<td>0.48</td>
<td>12.046</td>
<td>5.034</td>
</tr>
</tbody>
</table>

As for the M/HA pairs, since one would probably try to dominate, possibly the high prior knowledge student, while the other submits, collaboration is expected to be smooth, most especially if the two are already familiar with each other to begin with. This could be the reason for the M/HA pairs’ higher incidences of RR. The BH/HA and M/PA pairs’ RR’s and average performance scores were comparable. Their RR’s and average performance scores were in the middle.

A significant high negative relationship was only found between the M/HA pairs’ average performance score and RR (r = -0.731, p = 0.039) indicating that as their scores increased, their RR’s significantly decreased. The M/HA pairs had the lowest average performance score but with the highest average RR among the categories. This implies that their being familiar and comfortable with their partners and despite of having collaborated better did not warrant a better performance. Perhaps they just spent a great deal amount of time chatting or socializing. ANOVA and post hoc test results revealed that the RR’s of BH/PA and M/HA as well as M/HA and M/PA were statistically significant while others were not (see Table 1).

3.2 Determinism (DET)

The DET values of the BH/PA pairs were from low to average, the M/HA pairs had average to high DET, the BH/HA pairs had one high DET and majority were above the mean, the M/PA had one high DET, two low, and the rest were average (see Table 1 for high and low DET).

The BH/PA pairs had the least shared identical scanpaths and the M/HA pairs shared the most identical scanpaths. Possible explanation for this is the same in RR. High prior knowledge pairs might not feel the need to converge frequently but they still performed better nonetheless. It is assumed for pairs who are highly acquainted where one is a leader and the other a follower to have more matching scanpaths because most likely the follower would follow what the leader is aiming at on the screen. This could be the case of the M/HA pairs having better DET results compared to others, but their high DET turnout was the exact opposite of their performance. The M/HA’s average performance score was the lowest, possibly indicating that the pairs had engaged in more off-task behaviors. It could also be that the M/HA pairs had disagreed more frequently.

A significant negative high relationship only existed between the M/HA pairs’ performance score and DET (r = -0.789, p = 0.02), denoting that as more bugs were found, their matching scanpaths dropped. This implies that the M/HA pairs tried to work on their own but their scores did not improve. This could also be the case of more divergent episodes. The BH/PA pairs had the highest average score but with the lowest DET, possibly conveying that the BH/PA pairs collaborated less but their scores were still the highest among the categories. Their divergent episodes trying to build program understanding on their own could have contributed to their better performance.

The BH/HA pairs did not perform better as the BH/PA pairs. Though their average RR was the second highest among the categories, their average score was in the bottom two. This indicates that their being familiar and at ease with each other could have affected their performance. They might have spent more time on off-task situations or chatting rather than on the main task. The M/PA pairs average performance score was in the top two but their average RR was in the bottom two, signifying that their
being unfamiliar with each other caused them to engage in more divergent episodes and in doing so, their performance scores increased. ANOVA and post hoc tests showed that BH/HA and BH/PA, BH/PA and M/HA were statistically significant while others were not (see Table 1).

3.3 Entropy (ENTR)

The BH/PA pairs had the least complicated scanpaths as half of its ENTR values were low, whereas the M/HA pairs had the most complicated scanpaths because of its ENTR values which ranged from average to high. Refer to Table 1 for high and low ENTR value. The BH/PA pairs had the highest average score but with the lowest average ENTR among the categories. This could probably mean that their predictable but proven and tested debugging strategies, which implied more consistent or steadier scanpaths, could have contributed to their high performance scores. Their being unfamiliar was irrelevant since they still performed the best among the categories.

The M/HA pairs had the lowest average score but with the highest average ENTR. This could mean that they possibly did shotgun debugging which made their scanpaths the most complex since they just tended to look anywhere on the screen and this made their average performance score the lowest. Their being highly acquainted causing them to spend more on off-task behaviors could have contributed to their poor performance scores.

There were no significant correlations between the performance score and ENTR in all categories. However, ANOVA per program and post hoc tests showed significant differences between BH/HA and BH/PA, BH/PA and M/HA, and M/PA and M/HA. ANOVA overall revealed significant differences only between the BH/PA and M/HA pairs (see Table 1).

3.4 Laminarity (LAM)

Half of the BH/PA pairs’ LAM values were low and the rest were average. The BH/HA, M/HA, and M/PA pairs only had average to high LAM with the M/HA pairs having more high LAM compared to other categories. See Table 1 for high and low LAM values.

This could mean that the BH/PA pairs did not spend as much time as the other categories on certain programs or regions of code, and hence, they transitioned faster to other slides and were faster in terms of debugging. This is an indication of better program comprehension. This was confirmed in their average slide switches between the program specification and actual program and also their average fixation points and fixation duration, which were the lowest among the categories. Their average performance score was also the highest among the categories. Their degree of acquaintanceship was irrelevant in relationship to their performance scores.

The M/HA pairs tended to spend the most time on certain regions of the code but this did not translate to better scores since their average score was the lowest. It is possible that in those moments, they were just chatting and were not really concerned with finding the bugs in the programs. Their being acquainted might have caused their poor performance scores possibly because of too much spending in off-task behaviors and socializing. The BH/HA pairs despite being both skilled did not perform well as the BH/PA pairs. Their familiarity might have caused their weak performance focusing more on off-task behaviors and socialization. Their highest average fixation points and fixation duration as well as their average high LAM next to the M/HA pairs indicated that they might have socialized more.

The M/PA pairs performed well next to the BH/PA pairs and their average LAM value was the second lowest. This indicates that the M/PA pairs preferred to engage in more divergent episodes because of their unfamiliarity with their partners. In doing so, their performance scores improved. ANOVA and post hoc tests showed significant differences between BH/HA and BH/PA, BH/PA and M/HA, and M/PA and M/HA (see Table 1).

4. Summary and Conclusion

This paper compared the collaboration and performance of pairs consisting of two individuals who may have different or same level of prior knowledge and who may be highly or poorly acquainted given the task of program tracing and debugging. High-performing pairs who are poorly acquainted (BH/PA)
collaborated the least (low RR and DET) but performed the best. This confirmed the findings of prior studies that skilled strangers perform best. Mixed prior knowledge pairs who are highly acquainted (M/HA) collaborated the most (high RR and DET) but they performed weakly implying that their familiarity and being comfortable with each other did not warrant a better performance.

The BH/PA pairs’ least complicated scanpaths (low ENTR) translated to better performance. The M/HA pairs’ more complicated scanpaths (high ENTR) possibly indicated the use of trial-and-error debugging strategies, which resulted to a weak performance. The BH/PA pairs spent the least amount of time on certain regions of code (low LAM). This indicated better comprehension and better performance. The M/HA pairs had the highest LAM but with the lowest average performance score. High performing pairs who are highly acquainted (BH/HA), despite of being both skilled, did not perform well as the BH/PA pairs. Mixed prior knowledge pairs who are poorly acquainted (M/PA) performed well next to the BH/PA pairs but their LAM values were among the lowest suggesting that they preferred to engage more in divergent episodes which translated to better scores.

These preliminary findings confirmed that friendships or strong familiarity with partners could detract from performing well and are linked to reduced productivity because they have a tendency to spend less time focusing on the task and instead spend more time socializing. This also confirmed that friendship is irrelevant to better performance when groups are composed of highly skilled individuals. Further results of this study could help educators in teaching introductory computer programming where attrition rate is known to be high. Collaborative learning tasks such as pair programming could be strengthened by pairing students who would most likely to collaborate the most and perform at its best. Future work will look at their discourse data and triangulate it with the eye-tracking data to further validate the results of this study.

References


A Toolkit for Action: Translating Theory into Practice

Lucila CARVALHOa* & Pippa YEOMANb

aInstitute of Education, Massey University, New Zealand
bCentre for Research on Learning and Innovation, The University of Sydney, Australia

Abstract: Educators are increasingly being asked to promote 21st century skills in their digitally-enhanced classrooms, supporting students’ development of critical and creative thinking, problem-solving skills, and to learn to collaborate as productive members of an increasingly netwoked society. Educators, tasked with the oversight of increasingly active and digitally-enhanced learning spaces, are in need of practical methods and tools capable of supporting theoretically informed practices in design for learning. In this paper, our aim is to translate recent theoretical developments in the learning sciences into methods and tools for action, in order that the rich opportunities for learning that these spaces offer, do not go to waste. Central to this research, is the need for analytical tools that facilitate conversations about learning theories and design practices, between different members of increasingly diverse educational design teams. The paper introduces A Toolkit for Action, as a tool and a method for mediating conversations about designing for learning in the digital age.

Keywords: Design for Learning, Embodied Cognition, Distributed Cognition

1. Introduction

Teaching in the 21st Century requires that educators (re)design tasks that help students develop critical and creative thinking, problem-solving skills, and an ability to work as a productive member of a team—to learn to collaborate (Johnson et al, 2015). Schools across the globe are changing, as students are recast as active creators and not passive recipients of knowledge. These transformations occur not only in how people teach, but also in where they teach, as learning spaces are altered to accommodate different forms of innovation. Our research draws on theories of embodied cognition (Clark, 2010; Kirsch, 2013), entanglement (Hodder, 2012), the materiality of learning (Sørensen, 2009), and pattern languages (Alexander et al., 1977); as well as on methods from design anthropology (Gunn et al., 2013), design thinking (Brown and Katz, 2011), and the learning sciences (Carvalho and Goodyear, 2014; Carvalho et al., 2017). We argue that in order to develop successful computer-supported teaching and learning practices—in innovative learning spaces—we need to understand how productive learning activity emerges through the interplay of people, materials (both physical and digital) and task design. Educators tasked with the oversight of increasingly active and networked learning spaces, are in need of practical methods and tools capable of supporting theoretically informed design for learning. Our aim is to translate recent theoretical developments in the learning sciences into methods and tools for action, in order that the rich opportunities for learning that they offer, do not go to waste. Our research question focuses on: How do we facilitate productive conversations, about theory and design, between different members of increasingly diverse educational design teams? We introduce the Toolkit for Action—both a method and tool for mediating conversations in educational design teams. Produced primarily as a set of cards, the toolkit acts as a translation device (Bernstein, 2000) to facilitate the process of bringing educational theory into educational design practice. In the next section, we introduce ideas from design and sociology, explaining the need for analytical devices to mediate knowledge discourse in educational design. We then discuss how the toolkit enacts ideas in the Activity-Centred Analysis and Design framework (Carvalho and Goodyear, 2014), before exploring how learning theories, such as embodied and distributed cognition, shed light on the role of things in thought and action—in learning. We describe the methods and the design of the toolkit, and outline our vision for
its future development and how it supports knowledgeable action, in learning design and the orchestration of learning activity.

2. Bringing Ideas from Design & Sociology into Design for Learning

Design projects often bring people from varied disciplinary backgrounds together, such as architects, engineers, builders and clients. Each set of stakeholders has idiosyncratic ways of seeing and speaking about design. But for successful collaborations to emerge designers need to find common ground in which to articulate their ideas (Carvalho et al., 2009), in order to develop a shared understanding of what is to be achieved. In these scenarios, no single actor is likely to possess all the knowledge necessary to realize a design, and it is only through conversations about different design ideas that a team can achieve a shared understanding of the best way to tackle a specific problem (Kleinsmann et al, 2012; McDonnell 2009). As such, a team’s effective engagement in processes of knowledge sharing is crucial to successful collaboration. Conversational Turns is a method used by McDonnell (2009) in the analysis of design conversations by a team of architects and building-users, as they engaged in collaborative design. The method helps in tracing the progression of design as a collaboratively negotiated task. Sequences of interaction in design meetings can be examined, and episodes of knowledge sharing and knowledge integration can be identified in the conversational setting of design tasks. Both knowledge sharing and knowledge integration are interconnected processes, established by the turns taken in conversations. In Conversational Turns, a series of tentative movements are identified, starting with someone explicitly acknowledging a position or knowledge put forward by another person in the team, and then using this proposition to recommend or justify a design decision, which in turn incites an expert response or confirmation (Keinsmann et al, 2012; McDonnell, 2009). Similarly to the design field, educational design also often involves mixed arrangements that bring together stakeholders from diverse backgrounds – for example, educational managers, space planners, teachers, instructional designers, students. Likewise in design, all of these stakeholders might bring diverse ways of looking at an educational design task. Educational designers exchange ideas about the creation knowledge artefacts for learning – discussing, thinking about, planning the use of, and developing artefacts for educational practices related to the production, recontextualization, teaching and learning of knowledge. Building on Basil Bernstein’s code theory and on Pierre Bourdieu’s field theory, Legitimation Code Theory (LCT) (Maton, 2014) acknowledges knowledge as both: (a) socially constructed, within cultural and historical conditions, and (b) as something in its own right, something that may take different forms and have diverse effects on educational practices. Knowledge claims tend to vary according to social contexts, and knowledge practices reflect implicit ‘rules of the game’ that operate in (and are specific to) these, affecting and shaping the way knowledge is expressed or communicated (Maton, 2014). On that view, an educational team is likely to bring together people who all practice design – a space planner, a principal, and a maths teacher – who may have different underlying rules for knowledge practices in design. They will bring these to conversations, as their perspectives will be grounded in particular ways of seeing and valuing knowledge. Our Toolkit for Action was designed as a translator device—bringing educational theory into design conversations—to facilitate processes of knowledge sharing and knowledge integration, in the context of the work done by educational design teams. Before illustrating how the kit supports these knowledge conversations, we introduce ACAD, a second theoretical framing that highlights different ‘topic areas’ of productive design conversations.

2.1 The Activity-Centred Analysis and Design Framework (ACAD)

The ACAD framework sees learning as socially and physically situated, but also as powerfully shaped by epistemologies of learning (Goodyear and Carvalho, 2014). ACAD foregrounds activity—or what learners do—as central in any given learning situation; and learning situations are conceptualised as structural compositions of tools, tasks and people. Aspects of learning that are open to adjustment—the designable components—are referred to as set, epistemic and social designs. In making the distinction between what is open to alteration and what is not, ACAD describes the activity of learners as emergent. Goodyear’s earlier (1999) work on pedagogical frameworks and Alexander’s work on Pattern Languages (1977) reveals multiple scale levels at which design can influence those things that are open
to adjustment. Building on these ideas, Yeoman (2015) created a three-by-three wireframe (Table 1) where concepts from the ACAD framework simultaneously converge in a horizontal correspondence across the three dimensions (set, social and epistemic), as well as a vertical correspondence at three different scale levels (the detail or micro, the regional or meso, and the global or macro). Yeoman’s wireframe lays out a way in which ACAD, Pedagogical Framework and Pattern Languages may be brought together to connect observations of materials in use, with different dimensions of design, and in so doing, it helps designers to account for what constitutes good design for learning. The wireframe acts as the backdrop for the translation device (Bernstein, 2000) that helps designers navigate between theoretical concepts and their practical enactments. One of the ways the wireframe does this is by supporting the fine-grained analysis of one element (smartphone) of one dimension (set), as it is enrolled in emergent learning activity—without losing sight of elements of the other dimensions at multiple scale levels. Next, we introduce two theoretical perspectives—embodied and distributed cognition, arguing that the use of knowledge artefacts may facilitate educational design activity—to mediate conversations, as well as to encourage collaboration. These theories claim that material things help humans think differently, and provide a rationale for how cards might help in thinking about tools, tasks, and people in design for learning.

### Table 1: The ACAD wireframe

<table>
<thead>
<tr>
<th>SET</th>
<th>SOCIAL</th>
<th>EPISTEMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro</td>
<td>Artefacts, tools &amp; texts</td>
<td>Roles &amp; divisions of labour</td>
</tr>
<tr>
<td>Meso</td>
<td>Allocation &amp; use of space</td>
<td>Community</td>
</tr>
<tr>
<td>Macro</td>
<td>Buildings &amp; technology</td>
<td>Organisational forms</td>
</tr>
</tbody>
</table>

#### 3. Lessons from Embodied & Distributed Cognition: Using Cards to Think

Embodied cognition speaks of extension of minds beyond bodies, to include tools, symbols, artefacts that mediate interactions in the world (Clark, 2010), where “the concepts and beliefs we have about the world are grounded in our perceptual-action experience with things, and the more we have tool mediated experiences the more our understanding of the world is situated in the way we interact through tools” (Kirsh, 2013, p. 3:3). In other words, cognition grounds our behavior and is influenced by our perceptual system, where alignment between our actions and predictions that we make about the environment co-evolve (Markauskaite and Goodyear, 2017). Design thinking has embraced ways of understanding culture and context through learning by doing—where making is an integral part of the process, and prototyping is as an activity that speeds up the design process (Brown and Katz, 2011). Design thinking also has a participatory character, one in which bouncing ideas around and exchanging experiences are key. We argue that the toolkit encourages designers to participate in practices that are representational and embodied. What is more, that by their very nature—as knowledge artefacts—the cards support the distributed knowledge practices of interdisciplinary design teams. Distributed cognition (Hutchins, 2014) is not another descriptive theorization of how it is we come to know. Rather, it asserts that all cognition is distributed and emerges from distributed processes. As Hutchins (2014) reminds us: “wherever we find cognition, it will be possible to investigate how a process we call cognitive emerges from the interactions among elements in some system”(p. 36). This way of theorizing resonates well with our research, and helps explain why people find working with the help of cards, supports design for learning, where no single person can attend to or resolve the complex interplay of set, social and epistemic design, across multiple scale levels.

#### 4. Research Methods

The playful way in which we encourage people to work the cards does not rest on theory alone, but draws on methods used in design anthropology. Design anthropology is an emerging field that combines the forward orientation of design with the retrospective strengths of anthropology. Knowledge production in design anthropology differs from traditional ethnography in that their use of theory goes ‘beyond analysis and description to the generation of design concepts’ (Otto and Smith, 2013). Working
collaboratively, for extended phases, design anthropologists help to generate and refine concepts (Drazin, 2013), and explore different styles of knowing (Kilbourn, 2013) using non-textual tools (e.g. perceptual synthesis, experience juxtaposition and others). An array of artefacts, images, craft supplies and construction toys can be used in the joint resolution of a shared problem/challenge. Our Toolkit for Action is grounded on the analysis of participant observation, field notes, and video and photographic records of design activity, conversations and artefacts from five Workshops, where participants used the cards as a medium to engage in design conversations.

5. A Toolkit for Action: Bridging Conversations about Theories & Practices

5.1 An Analytical Tool for Mediating Conversations

Based on tools from design anthropology (Gunn et al., 2013), and building on the research of Carvalho (2010), Chatteur (2011) and Yeoman (2018, 2015), we created our Toolkit for Action as a way to facilitate the enactment of theoretically informed educational design practices. We envisaged the toolkit as a tool for teams—an aid to stimulate processes of knowledge sharing and knowledge integration (McDonell, 2009) by supporting educational designers reach a shared understanding of design for learning, through mediated conversations. In its current form, the kit has four main elements: (1) a set of cards, (2) a set of artefacts, (3) three learning scenarios, and (4) the ACAD wireframe. The full kit includes 93 cards, distributed across four dimensions of design, with each assigned a specific colour. Blue cards (Figure 1) are associated with theoretical concepts and high level philosophy (Goodyear, 1999). They are presented in two sub-types. The first displays the names of learning theories and their principal authors, and the second offers quotes selected to prompt reflection and discussion about learning. The remaining three sets of cards are associated with one of the three dimensions of design, as defined in ACAD (Carvalho and Goddyear, 2014). Green cards offer headline terms associated with set design (space, place artifacts, tools and texts), such as collaborative learning studio, pen and paper, laptop, and Learning Management System. Second order terms on the Green cards include key words describing qualities of the named element. For example, the first order heading smartphone is described, in second order headings, as (in)formal, dynamic, blended. By this we mean to convey that a smartphone can be considered a tool for learning in both formal and informal settings, that it supports a dynamic or decentralised form of learning activity, and it facilitates blended learning. Yellow cards offer headline terms associated with epistemic design (task design), basic task structure such as pace and mode of delivery, content selection, and assessment. The second order headings on these cards simply identify them as, structure. Yellow cards also refer to task type, such as problem solving or game playing. The second order headings on these cards suggest broader sets of instructional forms. For example, problem solving is described as indirect instruction, whereas game playing is described in terms of experiential instruction. Orange cards offer headline terms associated with social design (individuals, groups, roles, divisions of labour), such as team or assigned roles. The second order headings on the orange cards differentiate between social-identity (the next user), social-instruction (scripted roles) and social-responsibility (facilitator). The Blue cards were specially designed to initiate theoretical conversations and to stimulate processes of knowledge sharing (McDonell, 2009). They encourage people to share a particular viewpoint about a high level philosophy, and invite others to respond to that viewpoint by adding their own views to the mix. The Green, Yellow and Orange cards also encourage knowledge sharing in relation to epistemic, social and set design. Moreover, in combination with other artefacts, such as the wireframe, the cards encourage processes of knowledge integration by providing an opportunity to test how different components of each dimension (set, social and epistemic) correspond or conflict with design choices and constraints of design across various scale levels (macro, meso and micro). This type of knowledge integration is further supported through the use of artefacts to represent the alignment of individual knowledge to a shared-like orientation, for example with a mix of butcher’s paper, stationary tools, or other artefacts. Current artefacts in the toolkit include photographs of various learning environments, floorplans, and an assortment of 2D and 3D architectural representations of different modes of use. It also includes three learning scenarios, carefully scripted and based on previous research conducted in museum, university and school settings (Carvalho, 2010; Carvalho and Garduno-Freeman, 2016; Yeoman, 2015).
They offer educational designers authentic designs, which they can re-present using the cards. In this way, those unfamiliar with learning design or ACAD can focus on the dimensions of design, and the coherence or dissonance between dimensions and scale levels, without having to create an entirely new learning design in the process. In addition to these curated artefacts, the kit would be decidedly less generative without a healthy supply of various resources that encourage collaborative design activity, such as Post-it notes, blank Artefact Cards, markers and butcher’s paper.

Figure 1: Blue Cards (left) and Cards in Action (right)

5.2 A Method for Using the Toolkit for Action

The toolkit is to be used in three complementary, but independent stages, which can be quickly adapted to reflect the skills and needs of an educational design team by, for example, selecting a subset of cards, or creating customised cards to scaffold a specific brief. **Stage 1** is about Learning to Bring Theory into Practice, and is designed to scaffold the use of the cards. At this level, educational designers learn how to use the cards with the help of the ACAD wireframe. The focus is on creating shared consensus about the learning theory at play, before documenting a clear articulation of that theory, and tracing its correspondence or dissonance across dimensions of design and scale levels. **Stage 1** workshops are appropriate when working with inexperienced educational designers, with newly established teams, where time is limited, and in the early stages of learning environment (re)development or curriculum renewal. **Stage 2** is about Bringing Theory into Design Practice and is designed to give educational designers an opportunity to work on a design challenge that is specific to their current context. It could be the (re)design of a particular unit of study or a new learning space. Some team work is aspirational, whilst others work hard to meet the challenges of a tight design brief and or limited budget. Either way, participants are encouraged to start with the Blue cards so that their work is theoretically grounded. They are then free to engage with the other cards and artefacts in a way that meets their needs. In **Stage 2** the focus is on identifying those elements of each dimension that are both open to design, and within participants’ sphere of control to influence—with the express intention of increasing the coherence of the learning whole. **Stage 2** workshops are appropriate when working with existing teams or experienced educational designers on projects with personal relevance. **Stage 3** is about Refining the Design and Orchestration on the Fly and is designed to offer designers time to reflect on and improve an existing design. Participants in **Stage 3** workshops learn how to use the ACAD wireframe to analyse a particular moment of learning activity, or how the use of a given tool shapes learning activity. **Stage 3** focuses on iterative improvement and adaptive flexibility in future implementations of an existing course. **Stage 3** workshops are appropriate for educational designers wanting to reflect on and improve their current teaching and learning practice, and for teams tasked with post occupancy evaluations of new learning spaces.

6. Conclusion & Future Research

As innovative digitally-enhanced learning spaces become common place, a broader cross section of educational designers will be required to engage in increasingly complex design projects. Finding ways to facilitate communication amongst diverse educational design teams is critical, if successful collaborations are to emerge through shared understandings of how to tackle complex problems. This paper introduces a method and a tool to support educational designers conversations, in a way that brings learning theories into the design scene. Ideas from sociology, design thinking, embodied and distributed cognition ground the creation of the Toolkit for Action (Carvaho and Yeoman, forthcoming). Our future research goal is practical, to continue to analyse our data and fine tune a set of 100 cards that can be scaled down by thirds across dimensions of design, without loosing functionality, and to develop
further scaffolds to assist instructional designers in resolving design challenges across a number of areas, including environment design, assessment design and innovative social designs. To do this we will continue to systematically explore the use of the toolkit with teams of designers, gathering relevant data to help us refine its design, artefacts and methods for use. We anticipate that through this process, and in gathering feedback from participants we will transform the Toolkit for Action into a universal method and tool.

Acknowledgements

This work was supported by an Australian Research Council Laureate Fellowship (Grant FL100100203) and an Australian Research Council Discovery Grant (DP150104163).

References

Carvalho, L., & Garduno-Freeman, C., 2016. CmyView: Walking together apart. 10th International Conference on Networked Learning, 9-11 May, Lancaster, United Kingdom.
Exploratory Analysis of Discourses between Students Engaged in a Debugging Task

Ma. Mercedes T. RODRIGO
Ateneo de Manila University, Philippines
mrodrigo@ateneo.edu

Abstract: This paper determined if and how high-performing and low-performing students differed in the language that they used as they collaborated on a debugging task. 180 students worked in pairs to debug 12 small programs with known errors. Students were segregated into high and low achievement levels based on the number of bugs they found. Chat transcripts from the pairs were analyzed using the Linguistic Inquiry and Word Count (LIWC) software. We found that high- and low-performing students only varied in terms of their use of words that implied discrepancy and sadness.

Keywords: discourse analysis, pair programming, LIWC

1. Introduction

Pair programming is a collaborative activity in which two programmers work together on a single computer in order to develop one software artifact (Williams, et al., 2000). Studies of pair programming in both industry and higher education showed that pairs are able to produce higher quality code at a faster rate than individuals. Furthermore, pair programming improves programmer confidence and job satisfaction. These benefits have been shown to transcend physical distance. Partners who were geographically separate tended to perform as well as those who were co-located (Hanks, 2005).

In recent years, researchers have tried to arrive at a more nuanced understanding of the communication between partners and how the collaborative process benefits student learning outcomes (Rodriguez, et al., 2017). Pair programming is a skill in itself, one in which partners have to manage what problem they want to solve, how the problem decomposes into elements, and the order in which these elements should be addressed (Zieris & Prechelt, 2014). These studies are important because they codify optimal discourse patterns that can then be taught to and learned by future generations of programmers.

The analysis of discourse is the quantification of utterances, i.e. spoken or written language (Chi, 1997). It involves the tabulation or counting of different kinds of utterances and the drawing of relationships between them. Within educational contexts, it is used to understand what students know or feel (De Wever, et al., 2006). While literature is rich with examples of analyses of student discourse in disciplines from psychology (e.g. Robinson et al., 2013) to math (Zemel, et al., 2007) to computer science (Romero, et al., 2013), discourse analysis of communications between partners in pair programming contexts, though, is still somewhat limited.

This paper is an initial exploratory analysis of conversations between partners engaged in a tracing and debugging task. The goal of this paper is to determine if and how high-performing and low-performing students differed in the language that they used as they collaborated. We attempt to compare the extents to which high- and low-performing students express cognitive and affective processes in their language use. We chose to focus on cognitive processes because the process of debugging involves a comprehension of programming syntax and logic and the ability to reconcile these with the program’s intention. We also focused on affective processes because past research has shown relationships between emotions such as confusion (Lee, et al., 2011), curiosity and engagement (e.g. Bosch & D’Mello, 2014) and student success.
2. Methods

The data collected and analyzed for this paper were part of a nationwide study on the behaviors of pairs of students who were asked to debug 12 programs with known errors. Six universities voluntarily participated in the study: Ateneo De Davao University (ADDU), Ateneo de Naga University (ADNU), Ateneo de Manila University (ADMU), University of the Cordilleras (UC), University of San Carlos (USC), and University of South Eastern Philippines (USEP). The universities were distributed from the north to the south of the country in an attempt to get as much geographic representation as possible—UC was in Baguio, ADMU was in Manila, ADNU was in Bicol, USC was in Cebu, ADDU and USEP were in Mindanao. USEP was the only state university. All the others were private.

2.1 Participants and Test Conditions

Students aged 18-23 years old and were in their first to graduate year levels. They had taken the college-level programming courses. Ninety (90) pairs of participants composed of 112 males and 68 females. There were more males than females in the sample because there generally tended to be more males than females in information technology degree programs.

The pairs were divided into two test conditions, static (S) and dynamic (D). In the static condition, students used a custom-built slide viewer program to read the code and identify the location of the bugs. There was no need for them to correct these errors. In the dynamic condition, students located and corrected the bugs using an integrated development environment.

2.2 Data Collection and Preparation

Participants took a pre-test to establish their levels of programming proficiency. Paired students were encouraged to consult with each other using a chat program. Although they were seated together in the same room, they were spaced far enough to ensure that all communications with their partner were via chat only. Students were given one hour to debug all 12 programs. How much time they chose to spend on each program was left to them.

2.3 Student Achievement

For the pre-test, students were given one point for every correct answer. For both the static and dynamic debugging task submissions, students were given one point for every bug that they correctly identified. The students did not incur penalties for wrong answers or blanks. To classify students as high (H) or low (L) performers, we first took the mean of the number of bugs found. Student who found less than the mean were classified as low performers.

2.4 LIWC

Conversations between pairs of students tended to be informal, hence students used a mix of English, Filipino (the national language), and local languages (Ilocano in Baguio, Bicolano in Naga, and Cebuano in Cebu and Davao). We contracted native speakers of these languages who were also proficient in English to translate the discourses to English. None of these native speakers were translators by profession.

We segregated the chats by student and then used the text processing program Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, et al., 2015) software to analyze each student’s transcript. Given a block of text, LIWC counts the number of words and then computes for the percentages of words within that text that fall under 92 categories. Our analysis is limited to the summary variables (analytic thinking, clout, authentic, emotional) and the affective (affective processes, positive emotion, negative emotion, anxiety, anger, sadness) and process categories (cognitive processes, insight, causation, discrepancy, tentative, certainty, differentiation) (Pennebaker, Boyd, et al., 2015). Analytic thinking refers to the contributor’s level of formality and structure as opposed to informal, personal, narrative language. Clout reflects a level of expertise and authority as opposed to humility, tentativeness, or anxiety. Authentic implies honest, personal, and disclosing text.
as opposed to guarded or distanced language. Finally, Emotional Tone refers to positivity versus anxiety, sadness, or hostility.

3. Results

We grouped the students by condition (S or D) and achievement level (H or L), resulting into four groups: static, high-performing (SH); static, low-performing (SL); dynamic, high-performing (DH); dynamic low-performing (DL). Table 1 shows the number of students per group, by gender.

<table>
<thead>
<tr>
<th></th>
<th>SH</th>
<th>SL</th>
<th>DH</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>39</td>
<td>18</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Female</td>
<td>23</td>
<td>12</td>
<td>10</td>
<td>23</td>
</tr>
</tbody>
</table>

On average, students typed 231 words, with a standard deviation of 233. The shortest text was 1 word, the longest was 1,505.

We used RStudio (RStudio Team, 2015) to perform two-way Analyses of Variance (ANOVAs) to compare the LIWC results among these four groups. High- and low-performing students did not vary significantly in terms of number of words used, $F(1,176)=1.161$, $p=0.28$. However, there was a main effect in terms of condition, $F(1,176)=9.55$, $p<0.01$. A post hoc Tukey HSD showed that the students in the SH group contributed more words than those in the DH ($p=0.02$) and DL groups ($p=0.04$).

3.1 Summary Variables

For Analytic Thinking, there was no main effect for condition, $F(1,176)=1.344$, $p=0.25$, or achievement level, $F(1,176)=1.001$, $p=0.32$. The interaction between the two factors was marginally significant, $F(1,176)=2.725$, $p=0.10$. However, a post hoc Tukey HSD test showed no significant difference among groups.

In terms of Clout, there was no main effect for achievement level, $F(1,176)=1.103$, $p=0.30$, or the interaction, $F(1,176)=0.53$, $p=0.47$. There was a marginally significant main effect for condition, $F(1,176)=3.134$, $p=0.08$. Once again, though, a post hoc Tukey HSD test showed no significant difference among groups.

For Authentic and Emotional Tone neither the omnibus ANOVA results nor the Tukey HSD results revealed any significant differences.

3.2 Cognitive Processes

The omnibus ANOVAs and Tukey HSDs did not yield significant results for Cognitive processes, Tentative, Certainty, and Differentiation.

The occurrence of insight words was significantly different between conditions $F(1,176)=5.82$, $p=0.02$. Insight was not significantly different between achievement levels, $F(1,176)=0.01$, $p=0.93$, nor in the interaction between the two factors $F(1,176)=0.28$, $p=0.60$. A post hoc Tukey HSD test showed no significant difference among groups.

There were some significant differences found in the occurrence of words implying causation. There was a main effect in terms of condition, $F(1,176)=11.80$, $p<0.01$, but not for achievement nor for the interaction. The post hoc Tukey HSD showed significant differences between the SH and DH groups ($p=0.03$) and the SL and DH groups ($p=0.01$). In both cases, students in the DH group used more causation words than their counterparts in other groups.

In terms of Discrepancy, we found a statistically significant difference for achievement level, $F(1,176)=9.17$, $p<0.01$, and a marginally significant difference for condition, $F(1,176)=3.43$, $p=0.07$. The post hoc Tukey HSD only showed that the SH group tended to use more discrepancy words than those in the DL group ($p<0.01$).
3.3 Affective Processes

The omnibus ANOVAs and Tukey HSDs did not yield significant results for Affective processes, Positive emotion, Anxiety, or Anger.

Negative emotions were marginally significantly different between high- and low-performing students, F(1,176)=3.55, p=0.06. A Tukey HSD showed that high performing students tend to express marginally significantly more negative emotions than low-performing students (adjusted p=0.06). No other group comparisons were significant.

In terms of words expressing sadness, there was a main effect for condition F(1,176)=8.50, p<0.01, and the interaction, F(1,176)=8.66, p<0.01. A Tukey HSD showed that students in the dynamic condition expressed more sadness (adjusted p<0.01). It also showed that the DH group tended to express more sadness than DL group (p<0.01), the SH group (p=0.01), and SL group (p=0.02).

4. Discussion

The goal of this paper was to determine if and how high-performing and low-performing students differed in the language that they used as they collaborated. Based on the findings, high- and low-performing students only varied in terms of their use of words that implied causation, discrepancy, negative emotion, and sadness. High performing groups used more causation words, with sentences like “Because we need to answer everything, apparently.” and “I thought it depends on who was compiling.” High performing students also tended to point out discrepancies with sentences like, “The last semicolon in line 12 is unnecessary.” and “Then you should count line 22 too.” High performing students also tended to express negative emotions such as anger, anxiety, or sadness by saying, “What the f*** is 4” or “You’re stupid”. For the subcategory of sadness, high performers tended to apologize, “Super sorry for the hassle!” and admit confusion “I’m now lost in p09”.

Although the unit of analysis was the individual rather than the pair, these findings are, to some extent, consistent with prior literature on effective programming pairs. The group of Rodriguez et al. (2017) found that, generally speaking, more conversation leads to better outcomes. In specific, feedback as well as meta-comments like admitting mistakes and shortcomings, e.g. “I think I’ll need help for every number hahaha sorry.” boost learning overall.

To determine which among these five features—Word Count, Causation, Discrepancy, Negative Emotion, and Sadness—were truly predictive of student achievement, we tried to built predictive models using Weka 3.6 (Frank, Hall, & Witten, 2016). A linear regression that tried to predict the debugging scores of the students had only two features: Sadness and Discrepancy. The final model was:

\[
\text{Debugging Score} = 1.7334 \times \text{Sadness} + 0.6187 \times \text{Discrepancy} + 12.3877
\]

A ten-fold cross-validation yielded a correlation coefficient of 0.26. The model shows that Debugging Score rises with Sadness and Discrepancy words.

This model, though, has to be regarded with some caution. On average, only 0.21% of words across all students are classified as sad and the values of Sadness range from 0 to 4.17%. Similarly, on average, Discrepancy accounts for 1.98% of words. Values of Discrepancy range from 0 to 18.18%. These relatively small values imply that the large epsilon probably accounts for most of the correlation.

We therefore attempted to predict whether the students were high- or low-achievers using a J48 decision tree. The final model is as follows:

\[
\text{Discrepancy} \leq 1.82
\]
\[
| \text{Sadness} \leq 0.72: \text{L (85.94/33.94)} \\
| \text{Sadness} > 0.72: \text{H (7.08/1.08)}
\]

\[
\text{Discrepancy} > 1.82
\]
\[
| \text{Sadness} \leq 0.51: \text{H (74.82/18.82)} \\
| \text{Sadness} > 0.51
\]
\[
| \text{Sadness} \leq 0.87: \text{L (8.09/1.09)} \\
| \text{Sadness} > 0.87: \text{H (6.07/1.07)}
\]
Note that Sadness and Discrepancy still continued to be the most predictive features. A 10-fold cross-validation yielded a kappa of 0.24 and an accuracy of 61%. While the Kappa is considered just fair, the accuracy of the model is still better than the majority class, which is 56%.

5. Summary and Future Work

This paper was intended to be an initial analysis of discourse between pairs of programmers engaged in a debugging task. Its goal was to determine differences in cognitive and affective processes, as expressed through language, among high- and low-performing students. The study found that only the occurrence of words implying Sadness or Discrepancy differentiated high- and low-performers.

The absence of other statistically detectible differences may be attributable to a number of limitations. First, the translations captured the gist but not the full nuance of the students’ discourse. Students frequently made use of slang such as such “wew” to denote “wow”. When they say “joke!”, they acknowledge an error, rather than attempt humor. They used “gg” to mean either “good game” or “gulong-gulo”. The former is a figure of speech that denotes defeat while the latter literally translates to “extremely mixed up,” an expression of confusion.

Second, LIWC is not designed to take into account the technical nature of the task. Certain words such as “problems” and “odd” were regarded as discrepancy words. When taken in context, though, students were using the former to refer to the problem number (1 through 12) that they were currently solving or that they were proceeding to solve. They used the latter because one of the problems had to do with odd and even numbers.

LIWC also did not process emoticons. Students tended to pepper their chats with smileys and similar expressions. These can all be indicators of a variety of affective states including confusion, anxiety, anger, and others that might have a relationship with overall performance.

Moving forward, we plan to continue the analysis of this dataset in a number of ways. There are many published fine-grained approaches to hand-labeling each contribution of each participant to arrive at speech acts that might imply be telling of student ability, effort, partner cohesion, or maturity of understanding. To capture these phenomena or others of interest, coding schemes will have to be developed, based on literature. For example, early work by Roschelle and Teasly (1995) described several discourse events that lead to the development of joint problem space between partners. These events include turn taking, socially distributed productions, repairs, and narrations. Farris & Sengupta (2014) attempted to trace the development of computational thinking between pairs of programmers. In more recent work of Rodriguez et al (2017), researchers categorized dialog moves between pairs of collaborating programmers as statements, opinions, explicit instructions, acknowledgements, questions, answers, feedback, or off-task statements.

Intuition tells us that high- and low-performing students should differ in ways that manifest themselves in behavior. This analysis and others like it are attempts to discover these differences in order to arrive at learnable patterns that can then lead to a more holistic approach of educating future programmers.

Acknowledgements

I thank the Ateneo de Davao University, Ateneo de Naga University, University of San Carlos, and University of Southeastern Philippines for allowing us to conduct the eye-tracking experiment. Many thanks also to Angie Ceniza, Kristian Cordero, Joanna Feliz Cortez, Josephine dela Cruz, Jeffrel Hermias, Joshua Martinez, Thelma Palaoag, Yancy Vance Paredes, Japheth Duane Samaco, Celesamae Vicente. Lastly, thank you to the Private Education Assistance Committee of the Fund for Assistance to Private Education for the grant entitled “Analysis of Novice Programmer Tracing and Debugging Skills using Eye Tracking Data” and the Ateneo de Manila University’s Loyola Schools for the grant entitled “Building Higher Education’s Capacity to Conduct Eye-tracking Research using the Analysis of Novice Programmer Tracing and Debugging Skills as a Proof of Concept.”
References


Analyzing a Practical Implementation of Training Metacognition through Solving Mathematical Word Problems

Tama DUANGNAMOL*†, Thepchai SUPNITHI, Gun SRIJUNTONGSIRI & Mitsuru IKEDA

*School of Information, Computer, and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University, Thailand
†National Electronics and Computer Technology Center, Thailand

Abstract: The aim of this paper is to demonstrate an analysis of a practical implementation of Computer-Supported Multi-Reflective Learning Model via MWP (CRLEM) by developing a web application system, called MathReflect. MathReflect shows that it can encourage a learner to reflect on their thinking process and familiarize with utilization of metacognitive questioning by graphical representation and meta-level discussion stimulation integrated with delivering appropriate metacognitive questioning at the right time and events. Especially for this study, a questionnaire for classifying a learner who has gained Seed Skill to become a self-regulated learner (Q-L2SRL) is developed and integrated with the data collected from the system as the system evaluation method.

Keywords: Metacognition, self-regulated learner, math word problem, peer assessment, collaborative learning, reflective learning

1. Introduction

From my past experience as an educator and a mathematics teacher, solving mathematical word problems (MWP) is like a bitter pill for students. This is consistent with reports from many standard tests (e.g. TIMSS, PISA, etc.) that many students have difficulties in learning MWP solving. The main difficulty that students encounter in solving MWP is to construct a problem model of a context by making inferences from the text (Jacobse & Harskamp, 2009). It was revealed by Schoenfeld (1992) that it is because they rarely take the time to monitor and regulate the use of cognitive strategies. This causes them to skip or misinterpret information from the problem and choose inappropriate solutions. The skills to monitor and regulate the use of cognitive strategies are involved in metacognitive skill. We found that this is an advantage feature of MWP solving which could be utilized as a medium to train metacognition instead of using a real life problem, which is ill-defined/unstructured. Due to complexity/implicitness of metacognition combining with an unstructured problem, it might be quite complicated and could cause frustration in novice learners.

The goal of training metacognitive skill is to help learners to be comfortable with applying meta-level thinking on their cognitive process and become self-regulated learners who can automatically monitor and regulate their learning processes and be aware of their difficulties to achieve their tasks. To become a self-regulated learner, a skill in which a learner induces themselves to comprehend their own cognition is required. We call this skill, “Seed Skill to become a self-regulated learner”. However, learning or training metacognition is not a simple task due to its implicitness. Even so, according to many studies, metacognitive skills can be taught to students to improve their learning (Duangnamol, Supnithi, Suntisrivaraporn, & Ikeda, 2015). To reduce the difficulties of training metacognition, cognitive targets from thinking processes in working memory are shifted to observable thinking processes (Kayashima, Inaba & Mizoguchi, 2005). The observation found in (Akanda, 2013) showed that reflecting on self-cognition enables learners to link their professional development to practical outcomes and to refine/broaden the understanding of what counts as useful activity. A new and
promising research subject thus may be assessing the effects of computer environments, which combine cognitive content with metacognitive support. Such metacognitive support can be designed in several ways, for example by using intelligent tutoring systems, educational multimedia systems, virtual agents, metacognitive hints, and so on (Nakano, Hirashima, & Takeuchi, 2002; Jacobse & Harsskamp, 2009). In our research, we propose Computer-Supported Multi-Reflective Learning Model via MWP (CRLEM | k3:lem |). It is composed of three components: Interactive Metacognitive Q/A environment, Representation Format of thinking processes and Meta-level Collaborative Discussion platform. Then, we implement the system, called MathReflect, to see how CRLEM supports learners to use intrinsic comprehension of metacognitive questioning to acquire Seed Skill to become self-regulated learners.

2. MathReflect System

We have developed a web application system, MathReflect, as an environment for CRLEM. MathReflect can be accessed via this URL (http://mathreflect.com). Figure 1 shows the system architecture. The system is composed of three components and a database. Corresponding to CRLEM, the three components are Interactive Metacognitive Q/A module (ImQA module), Thinking Representation editor (TR editor), and Metacognitive Collaborative Discussion platform (MCD platform). ImQA module has Metacognitive-Responding Agent (MrA) to automatically deliver metacognitive questions and encouragement messages and to receive responses from a learner. TR editor is composed of two components: QAS constructing toolkit (QAS is a sequence of questions and answers to acquire information to accomplish the solution of MWP) and InDi composing toolkit (InDi is a diagram showing a flow of information and its source/reason to be composed for accomplishing the solution of MWP). MCD platform is composed of two components, which are Peer Inspection toolkit and Collaborative Chat toolkit. MrA works in relation with both TR editor and MCD platform. It is active throughout a learning session to stimulate metacognitive learning atmosphere. In the next section, the learning architecture of the system and its flow of teaching/learning are explained.

3. MathReflect User Interface

To demonstrate how MathReflect works in practice, we review each phase in the system by traversing on its UI. MrA always appears in the bottom left of the active window, as shown in Figure 2. The interface of MrA has a message display window to deliver metacognitive messages to a learner, a timer to show time duration that a learner has spent in the session, and a text-input box with an accept/send button to get feedbacks from a learner. The following subsections provide more detail and examples to demonstrate the use of the system in each phase.

3.1 Overview Phase

In the first phase, the Overview Phase, MetaQ’s are used with our Description of Advantage Use of Metacognition (DAUM) and an example of QAS/InDi for introducing a learner to metacognition to provide them with meta-understanding of MWP solving. DAUM is composed of short explanation/definition of the relevant skills of a self-regulated learner and examples of questions to activate/encourage those skills. This content is shown in the Introduction Page to a learner when they start the Activity session. MrA provides time for the learner to read and make an understanding of the usual practice.
content, before delivering a metacognitive massage/question, e.g., “Please state your aim for studying MWP solving”, to initiate the learner to realize their goal to studying MWP solving.

3.2 Practice Phase

In the Practice Phase, MetaQ’s are used when the learner solves MWP by constructing QAS/InDi. Applying MetaQ’s while constructing QAS/InDi helps them to shift their meta-level thinking to base-level thinking to foster them to acquire meta-understanding of MWP solving.

This MWP for figure 3 and 4:

A measure of a vertex angle of an isosceles triangle is 87 degree. What are the measures of the rest of the angles of this triangle? (Use algebra to solve the problem)

Figure 3 shows an example of QAS. QAS is a sequence of questions and answers to acquire information on how to accomplish the solution of a given MWP. In QAS Constructing Page, see Figure 2, there are QAS constructing toolkit and QAS constructing support hint. QAS constructing toolkit is for facilitating a learner to form QAS. It is composed of, Input box: a pair of text inputs (one for a question and the other for its answer), Add button: for adding input box, Delete Selection button: for deleting a selected input box, Clear button: for clearing the work space, Submit button: for submitting the finished QAS into the system, Check Answer button: for checking correctness of ongoing QAS before submission. The following are possible responses from the system when the learner click on Check Answer button: 1) there is inappropriate order of sequence, 2) there is insufficient information contained, 3) there is irrelevant question contained, 4) there are n-wrong answers, and 5) the solution is incomplete. QAS constructing support hint is provided in the left column of the QAS Constructing Page. It is used to scaffold the learner to familiarize with vocabulary utilization of self-questioning in solving MWP. The following are the hints we provide for a learner:

1. List of possible questions/answers: contains a list of questions/answers with some irrelevant questions/answers in random order.
2. Filter only relevant questions/answers: contains a list of relevant questions/answers in random order.
3. Make proper order of the list of questions/answers: contains a list of relevant questions/answers in proper order.

Figure 4 shows an example of InDi. It is a diagram showing a flow of information and its source/reason to be composed for accomplishing the solution of MWP. InDi is composed of Information Node (in a rectangle)—to show information required, Information Tag (in top of each

Figure 2. Web interface of MathReflect at QAS Constructing Page

Figure 3. An example of QAS

Figure 4. An example of InDi
Information node—to indicate the source of the information (there are six tag options: Goal, Sub-Goal, Given Information, Hidden Information, Result, and Others), Order Link (black arrow)—to show consecutive order in which the information used, Reason (in a dashed rectangle over certain Information nodes)—to indicate why information applied, and Sequential Link (red arrow)—to illustrate the result which needs information that is not consecutively linked. To composing InDi, the learner has a task to selecting appropriate Information Tags and Reasons from the provided lists for existing information to make InDi complete. Before submitting InDi, the learner can check its correctness by clicking Check button. The system can report his/her number of wrong tags/reasons via pop-up window.

3.3 Peer Support Phase

In the last phase, the Peer Support Phase, MetaQ’s are applied to make the learner reflect on the performance of others and what they had done during inspecting peers’ works to prepare them new vocabulary and information in collaborative discussion. This phase makes the learner gain awareness of self-improvement on MWP solving. This phase is composed of two sub-phases: Peer Inspection phase and Metacognitive Collaborative Discussion (MCD) phase.

3.4 Analyzing MathReflect

4.1 Sampling Procedure

In this study the scope of our subjects are grade-9 Thai students who are confused and do not recognize/realize their difficulties in solving MWP. This is to differentiate those who have gained improvement using MathReflect from those who are self-regulated learners. To sample the subjects for our study, first, a teacher gives a lecture on general knowledge of MWP solving using algebra in traditional method to students. Then, those students are screened by a MWP solving test. The students
who failed the MWP solving test are selected to take Meta-Understanding in MWP Solving Questionnaire (MUMSQ). The questions in MUMSQ are shown as follows:

(1) Why can’t you solve the problem?
   
   (…) I don’t know! (…) I have no idea! or Express your reason: ……………………

(2) What is difficult for you that makes you fail to solve MWP?
   
   (…) I don’t know! (…) I have no idea! or Express your reason: ……………………

The students who cannot express their reason are selected as the subjects in this study.

4.2 Procedure of Training Metacognition with MathReflect

To train metacognition with MathReflect, our training program has 6 periods. Each period has 90 minutes. In the first period, a teacher hands on the program syllabus to all subjects, then, the 90 minutes of the period is divided into 20 minutes for the Overview Phase, 20 minutes for the Practice Phase, 45 minutes for the Peer Support Phase, and the last 5 minutes for MWP-quiz. In the 2nd – 6th period, the 90 minutes of each period is divided into 12 minutes for the Overview Phase, 20 minutes for the Practice Phase, 45 minutes for the Peer Support Phase, and the last 13 minutes for MWP-quiz.

4.3 Analysis of Practical Implementation of MathReflect

In this section, we show how to analyze the change of a learner after studying MWP solving with MathReflect to indicate whether the learner can perform MetaQing skill to acquire:

1. Meta-understanding of MWP solving (MU) and

To analyze the change of a learner after studying MWP solving with MathReflect, MU is considered in 3 dimensions: self-understanding, task understanding, and process understanding. Consequently, AS is considered in the same dimensions. More precisely, each dimension is considered in sub-categories as follows:

- **Self-understanding** \[S\] is composed of 5 categories: (i) attitude in studying MWP solving, (ii) goal of studying MWP solving, (iii) motivation in studying MWP solving, (iv) self-restriction in studying MWP solving, and (v) background knowledge for studying MWP solving.

- **Task understanding** \[T\] is composed of 4 categories: (i) principle/structure of MWP solving, (ii) knowledge required for studying MWP solving, (iii) factors influencing the complexity of MWP solving, and (iv) application/benefit of studying MWP solving.

- **Process understanding** \[P\] is composed of 4 categories: (i) MWP solving process order, (ii) obstacles during solving MWP, (iii) timing in solving MWP, and (iv) concentration during solving MWP.

The following scenario is used to demonstrate how to analyze the change of a learner. Due to a privacy issue, we name a volunteer JJ. JJ is a grade-9 student of a school in Sisaket province, Thailand. He failed our MWP-test. However, he was willing to be a volunteer in our study. This may imply that he would like to improve himself on MWP solving. The selected scenes happened in Overview Phase 1st round (OP1), Peer Support Phase 1st round (PP1), and Overview Phase 3rd round (OP3).

(Overview Phase: 1st) after 5 minutes on the Introduction Page, MrA asked JJ, “How well do you understand what you read”. Then JJ answered, “I am not sure!”. MrA responded, “It’s ok just move on”, and delivered another statement, “Please state your aim for studying MWP solving MQ1”. JJ seemed confused. He then asked the teacher how to reply. The teacher suggested that JJ read the program syllabus. JJ answered, “To understand the fundamental strategy in solving MWP using algebra and be able to apply it in daily life A1”.

(Peer Support Phase: 1st) MrA suggested a discussion topic, “What are learning goals of the others?MQ2 How is it important to set up learning goal MQ3”. One member in JJ’s group, AA, asked in the discussion space, “What is your learning goal MQ4”. JJ replied, “MrA asked me a similar question, I answered MrA by using the learning objective from the syllabus A2”. The other member, ZZ, replied, “I want to master solving MWP. How about you AA A3”. AA replied, “I agree with both of you”. JJ raised a question to peers, “Why is it important to set up the learning goal?”. ZZ replied, “I think if we know our goal we can plan for it” A4. AA replied, “It helps me to focus my attention on the study A5”. JJ replied, “Oh.. I never thought about it before! I agree I agree A6”.

(Overview Phase: 3rd) MrA raised a similar statement again, “Please state your goal for learning MWP solving MQ5”. This time JJ replied, “I want to be able to master modeling MWP A7”.

208
The selected scenes demonstrate the analysis of the goal of studying MWP solving category in Self-understanding dimension, denoted as [S-ii]. The system encouraged JJ to [S-ii] by MQ1. By asking the teacher, he showed that he was confused. In PP1, JJ showed evidence confirming that he had no idea about [S-ii] at first, although, from the teacher’s hint he had a rough idea that following the learning objective could be set as a learning goal, as shown in A2. Moreover, the group members’ responses in A3-5 could enhance his understanding about [S-ii]. In OP3, he changed his answer to the same statement using his own opinion, A7, which reflected his understanding in [S-ii].

Besides collecting data from the system, we also use a Questionnaire for classifying a learner who has gained Seed Skill to become a Self-regulated learner (Q-L2SRL) as a part of the evaluation method for MathReflect. Q-L2SRL has been especially developed for our study. The items in Q-L2SRL were developed to cover all categories in all 3 dimensions of MU and AS. As a result, there are 26 items in Q-L2SRL. Q-L2SRL is shown in this link, https://goo.gl/forms/RjL9y867tvIFmrMR2.

From the case of JJ, in comparing between the beginning and the end of the program, JJ improvement in MWP solving can be significantly detected. His performance in MWP solving is increased, observed from MWP-quizzes. He can solve a seen problem or similar MWP more comfortably. From his log data collected from the system, there is ample evidence to show that JJ has gained meta-understanding in MWP solving. He expressed his opinion more often into his responses/answers in the late period of the program. Corresponding with Q-L2SRL, his score for MU and AS are 2 and 1.85 (from the full score of 3), respectively. This implies that he has gained the Seed Skill to become a self-regulated learner.

5. Conclusion and Future Works

By implementing MathReflect, it is shown that CRLEM has a potential to support a learner to gain Seed Skill to become a self-regulated learner. The system can stimulate the learner to get used to performing metacognitive questioning skill to acquire meta-understanding of MWP solving and awareness of self-improvement on solving MWP. The system works by encouraging a learner to reflect on their thinking process and familiarize themselves with utilization of metacognitive questioning by graphical representation and meta-level discussion stimulation integrated with delivering appropriate metacognitive questioning at the right time and events. In the future work, we would like to expand our subjects to English- and Japanese-speaking students. Moreover, we plan to generalize our learning model, CRLEM, to be independent of MWP solving.

Acknowledgements

This research is supported in part by JAIST-SIIT-NECTEC dual degree program, and Development and Promotion of Science and Technology Talents Project (DPST scholarship), Thailand.

References


Designing the EMBeRS Summer School: Connecting Stakeholders in Learning, Teaching and Research

Kate THOMPSONa, Antje DANIELSONb, David GOSSELINC, Simon KNIGHTd, Roberto MARTINEZ-MALDONADOd, Roderic PARNELLe, Deana PENNINGTONf, Julia SVOBODA-GOUVEAg, Shirley VINCENTb, & Penny WHEELERi

aSchool of Education and Professional Studies, Griffith University, Australia
bMIT Energy Initiative, Massachusetts Institute of Technology, USA
cSchool of Natural Resources, University of Nebraska-Lincoln, USA
dConnected Intelligence Centre, University of Technology Sydney, Australia
eSchool of Earth Sciences and Environmental Sustainability, Northern Arizona University, USA
fDepartment of Geological Sciences, University of Texas at El Paso, USA
gSchool of Arts and Sciences, Tufts University, USA
hCenter for Environmental Education Research, National Council for Science and the Environment, USA
iLearning and Teaching Centre, Australian Catholic University, Australia

Abstract: In this paper, we describe our research investigating design, teaching and learning aspects of the EMBeRS Summer School. In 2016, thirteen graduate Environmental Science students participated in a ten-day Summer School to learn about interdisciplinary approaches to researching socio-environmental systems. Using the Employing Model-Based Reasoning in Socio-Environmental Synthesis (EMBeRS) approach, students learned about wicked problems, team composition, systems thinking and modelling, stakeholder management, and communication. They applied this approach to their own research, as well as to a case study, in order to, ultimately, further the EMBeRS approach in their own institutions. Learning sciences researchers, environmental science instructors and learners collaborated in design, teaching, and learning during the 2016 Summer School in order to co-create and co-configure the tasks, social arrangements, and tools for learning, teaching and design. This paper identifies four examples of connections between the stakeholders (researchers, instructors and learners), the tools that facilitated the connection, and the implications for learning, teaching and design.

Keywords: interdisciplinary problem-solving, collaboration, learning, design

1. Introduction and Background

Learning, teaching, and design for learning have been conceptualised in terms of complex networks of learners, instructors, designers, and researchers, integrating physical and digital spaces (Howard & Thompson, 2016; Jacobson & Wilensky, 2006). To understand the relationships between design, teaching and learning, strong connections must be made between researchers and practitioners. This paper presents the application of a design-inquiry framework to analyse the Employing Model-Based Reasoning in Socio-Environmental Synthesis (EMBeRS) Summer School. It involved stakeholders in design, teaching and learning, and all contributed to the co-creation and co-configuration of the tasks, social arrangements, and tools for learning. Four examples of connections are presented: between researchers and instructors; instructors and learners; researchers and learners; and researchers, learners and instructors. The tools that facilitated each of these connections are discussed, and the implications for learning, teaching and design.

Synthesis, the act of integrating knowledge, data, methods, and perspectives in pursuit of a more comprehensive understanding, across disciplinary and professional boundaries is at the heart of addressing important socio-environmental issues. Many environmental science programs are functionally multidisciplinary and struggle to synthesize knowledge across disciplines (Vincent et al.,...
Researchers, designers, instructors, and learners require guidance on how to more effectively accomplish their interdisciplinary goals, yet there is little evidence-based advice to be given beyond ensuring quality communication (O’Rourke et al., 2013). Methods for sharing knowledge in groups have previously been provided for group settings in professional fields (Brown, Lindgaard & Biddle, 2011), but little has been explicitly developed for interdisciplinary teams of scientists (Pennington, 2011). In such teams, this knowledge is complex, must be conveyed to team members with basic training in that field, and needs to be connected to achieve research outcomes that are truly synergistic.

Core to the development of EMBeRS was understanding the design of several common techniques of problem-solving and adapting elements of each to a socio-scientific context. Idea generation must be conducted in a way that ensures each idea is explained and all members of the group understand (Pennington, 2011). Time must be purposefully allocated for team members to try to make connections with their own research and generate novel, synergistic models of the problem (Fiore & Schooler, 2004; Pennington et al., 2016). Building on research from experiential learning theory (Kolb, 1984) and creativity (Brophy, 1998), three features of successful synthesis were identified: the ability to externalize one’s own disciplinary knowledge; promotion of active listening and individual reflection; and iterating between divergent and convergent thinking activities.

Model-based reasoning (MBR) is based on the concept that when faced with a problem-solving task, humans reason by constructing an internal mental model of the situations, events, and processes that comprise the problem, and that external representations can be used to facilitate construction of a mental model (Nersessian, 2009). MBR provides a cognitive explanation for boundary objects (Star & Griesemer, 1989) as key components that link across expert perspectives (Pennington, 2010). Pennington et al. (2016) identified the key stages of interdisciplinary problem-solving for socio-environmental synthesis as: (1) identifying an appropriate research question; (2) agreeing on a shared vocabulary; (3) co-creating boundary negotiating objects; and (4) deploying tools for visualizing and combining data, with the aim of (5) producing a new, connected model of understanding. The product of this negotiation is a model of the system under inquiry. Individual scientists contribute data to the model, building on their initial conclusions and further discussing the relationships between this model and other connected research.

Understanding the relationships between the components in a system of learning and design helps us to better understand why a design is successful, repeatable or transferable. We draw on the Activity Centred Analysis and Design (ACAD) framework (Carvalho & Goodyear, 2014) to map learning systems and design so that the activity of the learner is placed at the centre of the design. Research on the implementation of the EMBeRS approach with undergraduate students (Thompson et al., 2016) demonstrated the importance of considering the connections between the design, implementation and outcomes in order to inform redesign. We combined key concepts from design based research (Sandoval, 2014) and the ACAD framework to: organize multiple analytic techniques applied to complex datasets; allow tasks to be compared across learning settings; and connect design and theoretical assumptions with specific design decisions. (Figure 1). The designed learning environment encompasses multiple components of the learning environment: the digital and physical learning environments, tools, resources, as well as the tasks and social arrangements. Learner activity refers to the observable aspects of learner behaviour: their social interactions, how they approach and work through tasks, and how they communicate in talk and through the generation of written or computer-generated representations. The activity of the instructor is also important. Learning outcomes refer to measurable changes in learners over time.

![Figure 1. The combined design approach](image)

Much of the recent discussion about multimodal data for learning (MMDL) has been reported in the context of multimodal learning analytics. In multimodal learning analytics, multiple types of data such as speech, text, handwriting, sketches, action and gesture, affective states, neurophysiological markers, and eye gaze (Blikstein & Worsley, 2016) are used to collect data about learner activity.
Research discusses how these data types can be connected, such as Thompson et al. (2013), and the importance of considering multiple dimensions of learner data to gain a more holistic understanding of learning activity (Blikstein & Worsley, 2016). Essential to considering MMDL is a way to organise, connect, and make decisions based on the results of analyses. Thompson et al. (2013) argue that the selection of data can be related to the ACAD framework.

2. Methods

The team implemented training activities during a two-week Summer School for PhD students, in July 2016, at the University of Texas at El Paso. The stakeholders included: seven instructors from the EMBeRS team (backgrounds in geological sciences, earth sciences, environmental science); guest instructors (specializing in systems thinking, stakeholder management); five researchers (backgrounds in science education, learning sciences, linguistics, learning analytics); and thirteen graduate students (six males, seven females). Graduate students were selected based on their disciplinary background (including environmental science and engineering; archaeology; bioengineering; urban management; ecosystem science and sustainability; agriculture and biological engineering; agricultural economics; water science and management; water resources), letter of recommendation from their advisor, the stage of their PhD, and their interest in interdisciplinary science.

The Summer School guided participants through lightly structured activities that employed the key phases of interdisciplinary problem-solving. At the end of each day, the group had explicit time for reflection on these activities, using the ACAD framework to guide student understanding of design choices, learner activity and learning outcomes, and each night they were asked to engage in individual reflections about their learning. In addition to the collaborative, discursive synthesis tasks, students were also given an individual, written synthesis task.

Participants were asked to complete an initial survey about their disciplinary background, educational experience and other background information. Design documents were prepared by members of the team, and their implementation recorded as the Summer School progressed. Audio recordings were collected, transcribed, and the discourse coded (using automated learning analytic techniques developed by team members) for convergence around ideas and language, and disciplinary knowledge. Video recordings were collected and the artefacts analyzed to identify the tools used. Interviews were conducted after the Summer School in order to obtain the participants’ perspectives on their gains in understanding and abilities and to evaluate the effectiveness of the different activities included in the school. Participants were asked to discuss what they learned, which activities were most helpful, how they intended to use their new knowledge and skills, and how the Summer School could be improved.

3. Results

The design of the Summer School and individual tasks was carried out by instructors and researchers over many months. The final design of the workshop was agreed upon, and transferred to a shared visual representation (Figure 2). Visualising the design of the Summer School allowed instructors to identify links could be made between individual tasks (e.g. Challenges of interdisciplinary work and Mock solicitation, Day 1), and repetition (e.g. Written reflections) and to manage tasks to be completed in students’ own time (e.g. Written reflections, Written synthesis). For researchers, visualising the complexity of the design of the Summer School helped to guide the research questions and data collection, and ensure that appropriate data was collected to answer key research questions.

As can be seen in Figure 2, on most days there was a morning session (e.g. Toolbox project, Day 2), an afternoon session (e.g. Simplify with frameworks, Day 3), and tasks to be completed during the evenings (e.g. Written synthesis, Days 3, 5 and 8). Some tasks were repeated (e.g. Written reflection each evening) and others involved guest presenters (e.g. Simulation activity, Day 4). Each activity was also mapped using the combined design framework (Figure 1), identifying elements of the epistemic, social and set design and the design and learning conjectures. Figures 2b and 2c shows the design of two of the tasks designed for Day 1. The combined design framework provided an important link between researchers and instructors. The framework was used as a tool to prompt discussion and negotiation of
meaning around key terms. After each discussion, researchers better understood the design and learning intentions of the instructors. Instructors were able to articulate the assumptions that they made about learning and teaching, and conduct design of the tasks that ensured that epistemic, social and set elements were considered. Every task was visualized using the framework, which provided researchers with detailed representations of designed tasks that can be compared in future analyses.

Figure 2. Design of the Summer School, 2016

Students were given multiple opportunities to practise the EMBeRS approach. Important features of the approach include active listening, and respect for different disciplinary approaches to solving problems. A strong culture of trust and a rhythm of communication developed between the instructors and the students. Multiple students wrote about trust particularly in their reflective tasks:

The culture that has been established by the group was intentionally designed by the organizers of this workshop, and is one that creates a high level of trust, knowledge sharing, and respect. I believe that respect is at the center of the cultural values... The high level of trust can only be established in a safe space for talking and sharing your knowledge, where every member is supported, rather than judged. (Samantha, Day 4 reflection)

This was most apparent during the tasks led by guest instructors, when it became obvious that the emergent practices of the group had not been communicated. This experience connected the instructors and learners in an unanticipated way.

The culture that the group developed by using the EMBeRS model to communicate our ideas and bring them into a common space was readily apparent today when we introduced other members ... to the group dynamic via Skype. Because these people weren’t present in the room, and had not experienced the culture... communication with them during question and answer period was more strained. ... In other words, the trust that we developed during the previous days of the workshop had not yet developed. (Sandy, Day 4 reflection)

Following this reflection, time was devoted to articulating the co-constructed group practices and the Summer School culture explicitly. Briefing of subsequent guests included introductions, a slower pace, and the provision of time for connections to be made in the co-creation of a shared model of understanding. It was empowering for the students to articulate and encourage these practices.
Researchers connected with learners in ways separated from the instructors. This was done through interviews at the end of the Summer School, and also tasks that students were asked to complete in the evenings (written reflections and synthesis tasks). During the interviews, most participants reported that they: learned skills to participate in and lead interdisciplinary/transdisciplinary teams (77%); gained an enhanced understanding of interdisciplinary/transdisciplinary research processes (69%); gained understanding of multiple perspectives/disciplines (62%), and learned to integrate disciplinary knowledge and methods using interdisciplinary modeling tools (54%). In a post-program survey asking participants to rate the effectiveness of each activity, almost all rated all the activities as of very high or high value. In addition to the written reflections, on three evenings, students were asked to write a synthesis of three articles, which had a shared theme (the water-food-energy nexus), each from a different disciplinary perspective. The students were first asked to draft a synthesis on Day 3 of the workshop, with opportunities on days 5 and 8 to redraft. While the interdisciplinary synthesis practices that were developed through discussion and co-creation of artefacts during the Summer School are important for an environmental scientist, the skills to synthesise and communicate different disciplinary knowledge in writing are also essential. The learners had the opportunity to engage in this practice, and to observe how their ability to connect disciplinary knowledge developed as they learned the group skills in parallel. The researchers analysed these syntheses with respect to their inclusion of topics or themes from the sources, intra- and inter-textual synthesis, evaluation, and sourcing (which articles were explicitly referred to). This analysis was conducted across all available drafts, to better understand the evolution of the synthesis over time. Across the texts produced, clear differences could be identified between students and over time, with students varying in the number of sub-topics or idea units expressed, the sourcing of these from the three documents, and their evaluation towards a particular conclusion.

The ACAD framework was used to guide group reflection at the end of each day. Students were asked to identify learning outcomes, and researchers suggested additional outcomes as relevant. Students then identified their activity, and the researcher outlined the design. The intention was for students to understand the purpose of the tasks they had participated in, for them to ask questions, and for them to make connections between what they were doing and what they were learning. For example, the overarching aim of all the tasks designed for Day 1 (Figure 2) was for students to gain experience in enacting the EMBeRS approach to solving problems. Learning outcomes were identified, including abilities to simplify thoughts about own research, communicate with non-experts, learn about different ways of representing, [develop] social capital, [identify] social implications, see interactions between research, and [realise] different programs experience with representatives. Only a subset of these identified learning outcomes aligned with those of the instructors. The guided reflection was beneficial for both instructors and students. For students, it scaffolded the connection of tasks within a day, or between days, with the overall learning goals. The process also allowed the instructors to reflect on whether the design and learning intentions of each day were met, and to make adjustments to the design of the tasks on subsequent days to ensure that any misalignment was corrected.

4. Discussion and Conclusions: Implications for design, research and learning

Many changes were made for the 2017 Summer School based on the analysis presented above. Two main changes reflect the importance of designing for the co-creation of an environment for learning in which instructors, researchers, and learners can connect, trust and build a collaborative culture as well as models of understanding. The timing of the Summer School was chosen to ensure that more of the instructors could be present in person, and the ‘share your research’ task on the first day was extended to ensure that all students have the opportunity to work together. This tests the design conjecture: *more time will enable the culture to be co-created and co-configured.* The ultimate aim of the research is to generate new insights into effective synthesis practices. These insights will enable synthesis decision-makers (by which we mean research team leaders, learners, instructors, and program designers) to make informed decisions about designing and engaging in synthesis activities. The multimodal dataset captured the activity of learners, over time, as participants learned to identify and represent their own disciplinary knowledge; collaborate in an interdisciplinary team; and allow a shared problem model to emerge. Further analysis of the dataset is focused on identifying evidence of
disciplinary knowledge, interdisciplinary knowledge (the shared 'language'), and collaboration, and relating these to the design of tasks and instructional practices.

Given the complexity of the design of the EMBeRS Summer School, there were numerous learning objectives related to individual tasks, as well as the Summer School overall. One of the implications of using the ACAD framework is the importance of observing the co-creation and co-configuration of learning. Learners were given access to the design intentions every day (through the guided reflections), and developed relationships with the researchers and the instructors. They became important stakeholders in their own learning, and had significant power in that relationship. A follow-up survey has revealed that many of the participants have applied what they learned during the Summer School to planning the next stages of their dissertation with their advisors. They have also been using elements of the approach in professional settings including the design of workshops, presentations, and other interdisciplinary research. The tools used (the ACAD framework) as well as the social relationships (with researchers and instructors), and the development of a shared culture, were as important as the designed tasks in enabling these students to co-create and co-configure their learning.

References


Why Learners Fail in MOOCs? Investigating the Interplay of Online Academic Hardiness and Learning Engagement among MOOCs Learners

Tonny Meng-Lun KUOa*, Chin-Chung TSAIb & Jyun-Cheng WANGc

a, c Institute of Service Science, National Tsing Hua University, Taiwan
b Program of Learning Sciences, National Taiwan Normal University, Taiwan
*kuo@mx.nthu.edu.tw

Abstract: Although the proliferation of Massive Open Online Courses (MOOCs) has created highly interactive learning environment for higher education, the low completion rate (or high dropout rate) deteriorates the development of MOOC courses. This study explores why learners are not highly engaged in MOOCs from the perspective of academic hardiness and learning engagement. The interplay of online academic hardiness and online learning engagement is mapped through both structural equation model and predictive model. Our explanatory and predictive analysis found that commitment is the most important factor of online academic hardiness, significantly influencing learning engagement. Moreover, the role of challenge contributes much to cognitive and emotional engagement. Other interesting findings and instructional implication will be discussed in the paper.

Keywords: MOOCs, online academic hardiness, online learning engagement, learning analytics

1. Introduction

Even though the advancement of MOOCs was considered as disruptive innovative technologies that transform the landscape of postsecondary teaching and learning (Christensen & Weise, 2014), the massive enrollment lead to low completion rate (or high dropout rate) in MOOCs which deteriorates the development of MOOCs (Freitas, Morgan, & Gibson, 2015; Hew & Cheung, 2014; Jordan, 2014; Perna et. al, 2014). In view of low engagement of MOOCs, several researchers and practitioners are dedicated to investigating why learners are not engaged in MOOCs learning environment, such as student factors, course/program factors, and environmental factors (Carnoy et. al, 2012; Hart, 2012; Hew & Cheung, 2014; Lee & Choi, 2011).

Although many works have been devoted to student factors, we found that academic hardiness, an important psychological concept in the learning process, are missing in the current literature. The relationship between the academic hardiness and learners’ online learning engagement are unclear in the MOOCs context. Considering that taking an online course requires a great amount of self-regulation and efforts, we argue that academic hardiness could be an important construct in open online environment.

1.1 Academic Hardiness

Derived from Kobasa (1979)’s hardiness theory, academic hardiness is a useful framework to understand why certain people are more willing to engage in challenging work and cope with stressful jobs (Kobasa, 1979). Previous studies found that academic hardiness plays an important role in learning process, such as anxiety, self-efficacy, and academic performance (Benishek & Lopez, 2001; Wang & Tsai, 2016).

The academic hardiness is composed of three distinct cognitive processes: Commitment, Control, and Challenge (Benishek & Lopez, 2001; Benishek et al., 2005). Commitment refers to the dedication...
of student’s reported willingness on specific goals or context to make sense of the meaning of purpose in life. Control is defined as learners’ belief that they possess to achieve personal desirable educational goals through efforts and effective emotional self-regulation. Challenge is the students’ purposeful efforts that students believe to be important to achieve higher levels of goals in terms of more demanding tasks or experiences.

1.2 Learning Engagement

Students’ learning engagement refers to the quality of effort students make to perform well and achieve desired outcomes, including behavioral, emotional and cognitive engagement (Henrie, Halverson, & Graham, 2015; Sun & Rueda, 2012; Vaughan, 2010). Behavioral engagement refers to the observable behaviors necessary to academic success, such as attendance, participation, and homework completion in class. Emotional engagement includes both feelings learners possess about their learning experiences, such as interest, enthusiasm, interest, enjoyment, vitality, frustration, or boredom, and their social connection with others at school. Cognitive engagement is the concentrated effort learners give to effectively understand what is being taught, including self-regulation and metacognitive behaviors (Fredricks et al., 2004; Skinner, Furrer, Marchand, & Kinderman, 2008).

Even though online learning engagement in MOOCs courses are widely explored, most measures of learning engagement in MOOCs employed behavioral engagement (such as count data) out of convenience purpose, ignoring the important role of motivational factors such as emotional and cognitive engagement. Moreover, learning engagement accounts a great amount of learning performance in MOOCs, by analyzing factors that led to high/low engagement depicts a better picture for MOOCs stakeholders, such as researchers, practitioners, and platform designers. In view of the indispensable role of learning engagement in MOOCs, the current study aims to understand how online academic hardiness influences learning engagement.

1.3 Purposes of the Study

This study explores why some learners are more engaged in MOOCs courses while some are not from the perspective of academic hardiness and learning engagement. In particular, from the structural relationship perspective, we would like to investigate the interplay of online academic hardiness and learning engagement in MOOCs environment in Taiwan. From the predictive perspective, we also examine significant factors that contributed much to online learning engagement through building predictive model using data mining techniques (Shmueli & Koppius, 2011). To distinguish explanatory and predictive goals, the following questions drive the study.

1. What is the relationship of online academic hardiness and learning engagement?
2. What are significant predictive variables for online learning engagement? How does the predictive model perform?

2. Method

2.1 Participants and Context of the Study

Previous studies on MOOCs take advantages of log data, interviews, surveys and web content analysis. However, to better understand the learners’ psychological attributes, this study employs validate survey questionnaire to conduct the study. We partnered with one of the popular MOOCs platform in Taiwan and distributed the survey at the platform at the end of the courses in 2017 March and April. Each participant taking one of the courses would take less than ten minutes to carry out the survey.

To better represent the context of the MOOCs, six courses were selected from different disciplines including life science (Eco-system and Global Changes; Systems Neuroscience), computer science (Introduction to Data Structure; Introduction to IoT), and business and management (Topics on Investment; Understanding & Rethinking Media). The courses have been run on a yearly basis and the course designs are similar in nature. All the courses are video-based learning MOOCs (xMOOCs), namely instructional videos are the main content of the MOOCs to identify important concepts. The
supported learning activities including discussion board, weekly quizzes, and assignments are to help student better understand the course content. The midterm/final examinations are to evaluate the effectiveness of the studies.

Of all the 1266 online survey distributed, 665 respondents successfully completed the survey (completion rate 52.52%). To detect anomalies data of the survey questionnaires, we used Rasch fit statistics from Item Response Theory (Chien et al., 2007) and finally excluded 57 responses of which outfit and infit values are above 2. In total, 608 observations constituted the study. Among the 608 observations, male learners represented 46.7%, female learners represented 53.0%, and other gender represented 0.3%. The average age was 23.93 [standard deviation (SD) = 5.78], which indicate that most leaners were undergraduate or graduate students. Most of the participants obtained a bachelor degree (74.4%), while 21.4% reported a master degree, 3.5% received high school degree and 0.8% obtained doctoral degree respectively. Of all the learners, more than half students (43.25%) had experiences of online courses and had completed 1-2 courses, 12.5% of the respondents had completed at least three online courses, 29.44% of the respondents had experiences of online course but failed to complete, and 14.8% of the respondents took the online course the first time.

2.2 Data collection and Analysis

The data used in this study were obtained mainly from convenience samplings from online survey website in Chinese. The instruments of the study were adopted from existing validated scales. Both scales used 5-point Likert rating (5 = strongly agree, 4 = agree, 3 = neither agree nor disagree, 2 = disagree and 1 = strongly disagree). The online academic hardiness scale (OAH) was adopted from Creed, Conlon, & Dhalliwal (2013) and Wang & Tsai (2016) while the learning engagement scale (LE) was adopted and modified from Sun & Rueda (2012).

To gain the expert validity, the questionnaire was first drafted and sent to two academic professors and five students for internal interview. Based on their comments and suggestion, we revised the wording and items to improve the scale quality. All the data were transformed and coded using RStudio software (Version 1.0.136). The R packages included psych, corplot, lavaan, semPlot, TAM, sirt, randomForest, and caret. We used TAM and sirt packages to formulate the Rasch model for detecting anomaly data; psych and corplot packages are to calculate the reliability and correlation matrix; lavaan and semPlot packages are used for modeling structural equation modeling (SEM).

3. Results

3.1 Test of the Research Model

We used lavaan and semPlot packages to calculate the overall fit and explanatory power of the SEM model in order to conform that relationship among constructs as expected in research model. Based on Hu & Bentle (1999)’s criteria, the chi-square ($\chi^2$=236) statistic is 955.3 (p<0.001), the comparative fit index (CFI) is 0.9 (should ≥ 0.9), the Tucker-Lewis Index (TLI) is 0.9, the root-mean-square error of approximation (RMSEA) is 0.07 (should ≤ 0.06), and the standardized root-mean-square residual (SRMR) is 0.07 (should ≤ 0.8). The above statistics suggests that our research model provide good model fit.

The results show that most of the hypothesis are supported. Commitment is found to be most important construct for online learning engagement, indicating that learners with higher commitment would be more engaged in the online learning environment. Challenge is also found to be an important variable to emotional and cognitive engagement. The path coefficients depict that challenge significantly contributes to emotional and cognitive engagement, implying that in order to increase learners’ emotional and cognitive engagement, creating a challenging environment might be a good way. Contrary to our hypothesis, control is found to have a little impact on behavioral and cognitive engagement and only possess a slight impact (path coefficient = 0.09, p<0.05).

3.2 Analysis of Predictive Model
In order to evaluate the predictive accuracy and predictive power, random forest data mining algorithm was used to build the model. The online learning engagement was transformed into binary category in terms of high and low engagement based on average engagement. We randomly partitioned data into training set (426 observations) and validation set (176 observations). Moreover, we used mean decrease accuracy (MDA) to better understand important predictors of the three models. Table 1 lists the top five predictors that contribute the most to the model.

Consistent with relationship modeling, we found that commitment the most significant factor to the three models. In the behavioral engagement model, learners’ online learning experiences play an important role in the model, while cognitive engagement, challenge, and emotional engagement are also important variables. As for emotional engagement model, cognitive engagement, behavioral engagement and challenge are important variables; interestingly, learners’ education to our surprise has a great impact on the emotional engagement model, indicating that learners with higher education would be more emotionally engaged in the online courses. In cognitive engagement model, learners’ age could be considered significant factors that improve learning. This finding is reasonable because learners’ cognitive ability might be highly related to their age.

Table 1: Important variables of three predictive model using mean decrease accuracy.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Mean Decrease Accuracy</th>
<th>Behavioral Engagement</th>
<th>Emotional Engagement</th>
<th>Cognitive Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>COM (32.17)</td>
<td>COM (61.93)</td>
<td>COM (27.47)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CE (15.31)</td>
<td>CE (60.10)</td>
<td>EE (23.47)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CHA (11.62)</td>
<td>BE (21.28)</td>
<td>BE (12.77)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>OCE (9.55)</td>
<td>CHA (20.06)</td>
<td>CHA (10.35)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>EE (8.12)</td>
<td>EDU (16.02)</td>
<td>AGE (10.08)</td>
<td></td>
</tr>
</tbody>
</table>

Note: COM = commitment; CHA = challenge; CON= control; BE = behavioral engagement; EE = emotional engagement; CE = cognitive engagement; OCE = online course experiences; EDU = education; AGE = age of the learner

4. Instructional Implications and Conclusion

The present study investigates the relationship between online academic hardiness and online learning engagement through both structural equation model and predictive model. Based on the previous findings (Fredricks et al., 2004; Sun & Rueda, 2012; Wang & Tsai, 2016), our study improves the understanding of the impact of online academic hardiness on online learning engagement (see table 2 for the summary of the research model). Our data and analysis indicate that:

- **Commitment** is the most significant factor to learning engagement either in explanatory or predictive model, meaning that increasing learners’ commitment in online learning environment may improve learners’ overall engagement. Based on this finding, we suggest that online instructors and teaching assistants on MOOCs should pay attention to learners’ online leaning commitment by implementing strategies that encourage learning commitment. For example, instructors could perform their hardiness and efforts in instructional videos to trigger learners’ commitment (Wang & Tsai, 2016). Additionally, platform designers, psychologists, and learning scientists could work together to brainstorm plausible ways to improve students’ commitment through the design of the platform and activities. By emphasis on peer-to-peer interaction and deliberate practice for commitment, learners would be more engaged in online learning environment.

- **Challenge** plays an important role in emotional and cognitive engagement, which implies that creating a challenging learning environment makes learners cognitive and emotional more involved in MOOCs learning process. Instructors should avoid over-simplifying the learning content for improving understanding in MOOCs; rather, they should design problems or activities to arouse cognitive conflicts for better cognitive learning. In designing such activity, conflict map could be used to design problems that foster scientific learning (Tsai, 2000; Tsai & Chang, 2005). Moreover, asking challenging problems before the instructional video as driving questions may arouse students’ thinking and attention. We suggest MOOCs content designers evaluate the level
of learning tasks based on target audience’s proficiency to increase emotional and cognitive engagement.

- Our predictive analytics found that learners’ prior online learning experiences are also important to behavioral engagement, which reflects the constructivism view of learning. Valuing learners’ prior knowledge and design appropriate learning content is of great importance in online learning environment (Tsai, 1998). Even though learners on MOOCs are claimed from diverse variety, we suggest MOOCs instructors analyze learners’ prior knowledge in advance for better design of engagement and effectiveness. For instance, content designers could encourage learners to (1) share prior experiences that links to the learning content in discussion board or (2) reflect their personal experiences as reflective journals to make more connection between knowledge and experiences.

Table 2: Summary and path estimate of the study.

<table>
<thead>
<tr>
<th></th>
<th>Behavioral engagement</th>
<th>Emotional engagement</th>
<th>Cognitive engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commitment</td>
<td>0.59 *** (+)</td>
<td>0.32 *** (+)</td>
<td>0.33 *** (+)</td>
</tr>
<tr>
<td>Challenge</td>
<td>N.S.</td>
<td>0.85 *** (+)</td>
<td>0.85 *** (+)</td>
</tr>
<tr>
<td>Control</td>
<td>N.S.</td>
<td>0.09* (+)</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

Note: *** $p < 0.00$, ** $p < 0.01$, * $p < 0.05$; N.S. = not significant

Our study presents the interplay of online learning hardiness and online learning engagement in a more holistic way to better understand why some learners might fail in MOOCs. However, three limitations should be stated for further improvement. First, we used convenience sampling in the current study, the data may not be able to reflect all the phenomenon in MOOCs. Future studies could employ more rigorous techniques of MOOCs data collection, such as stratified sampling or experiment to improve the data quality. Second, the predictive model of the validation set might be overfitting in predicting future engagement. More robustness testing or data mining techniques could be considered in building future engagement predictive models. Third, the current study takes advantages of self-reported data to map the relationship of academic hardiness and engagement. Future studies could focus on more authentic data sources, such as eye-tracking, brainwave or Galvanic skin response (GSR) to measure learners’ engagement. We sincerely hope that the current study benefits both academia and practitioners by incorporating path analysis and data mining techniques and there would be more studies on online academic hardiness and learning engagement.

Acknowledgements

This project is partially funded by Ministry of Science and Technology of Taiwan. The authors would like to thank Dr. Jian-Wei Tzeng and Prof. Cheng-Te Chen for their dedication to this study. We also like to thank three anonymous paper reviewers for their informative comments.

References


Toddlers Testing DDMM: Evaluation Results and Ideas towards Creating Better Learning Environments for Small Children

Christine STEINMEIER* & Dominic BECKING
Bielefeld University of Applied Sciences, Germany
*esteinmeier@fh-bielefeld.de

Abstract: Most musical instruments provide children with a rather complex user interface and unfamiliar possibilities of interaction, which do not appear in their daily experience of the world. This complicates the process of creating harmonic melodies and may lead to dissonance and frustration. Thus, the main goal of our research project „Drum-Dance-Music-Machine“ is the development of a toolset for early childhood music education, that facilitates a low-threshold access to collaborative music making. For this purpose, a prototypical new instrument has been constructed, offering the simple interface of a drum, but producing notes and melodies. Furthermore, with this system it is possible to create music by dancing and changing its parameters through movement. Amongst others, this aspect can be used to explain and clarify the meaning of abstract musical terms. During the ongoing development and after further enhancements we tested the system in two phases in order to ascertain our predefined goals. Amongst others, the children were asked to give their mind on their rating of this new instrument and tell if it is easy or difficult to create music with. Furthermore, we collected subjective assumptions from parents and educators, test observations and objective measurement data, combined them and analysed the results in terms of various aspects to verify the suitability of the system for use in early education.

Keywords: music education, early childhood education, evaluation, visual feedback

1. Introduction

1.1 Motivation

There is evidence that musical activities can bring many benefits on children's personal, intellectual and social development. (Hallam, 2016) However, most instruments traditionally used in music education do not have suitable affordances to provide very young children with the possibilities of fully exploring collaborative music making. In contrast, familiar user interfaces can ease this and motivate children to create music. The simplest interaction even very young toddlers can use is non-target-oriented pounding on surfaces and thus percussive instruments are particularly suitable as a low-threshold entry into music making. (Reifinger, 2006) They are designed for the generation of sounds with less or no defined pitch though and can, in principle, only be used for the exploration of rhythms. Manipulating tones and examining melodies, however, arises a special feeling of success and can thereby motivate to continue in making music. (Koelsch et al., 2003) So the use of a new musical instrument, which offers intuitive input possibilities and at the same time provides an extensive musical sound experience, would be more suited than classical instruments. Since particularly the making of music in groups is challenging in terms of consonance, such an instrument should also take this problem into account. In this connection, the understanding and correct use of musical terms and concepts plays an important role, since they are necessary for communicating about the common procedure. Instead of illustrating these abstract concepts with the aid of theoretical explanations, it would make more sense to experience them through direct application while making music. For example, the musical terms could be transferred into the colloquial language and linked with corresponding gestures.
1.2 DDMM Prototype

Based on the above-mentioned motivation, a system, which is supposed to close this gap, was designed within the framework of the interdisciplinary research project "Drum-Dance-Music-Machine" (short: DDMM). In 2015 we developed the first prototype, which turns a PC, a pair of electronic drum pads and a camera into an instrument that can be used by small groups of children. (Becking, Steinmeier & Kroos, 2016) The version we presented describes a modular software system, consisting of several VST plugins, which can be operated and connected in a host application. The main component of this construct is the plugin called “Music”, which is, among other things, responsible for the creation of a dynamic composition. Two other plugins receive and analyse data from connected Drumpads and a camera to generate different parameters. So far playing with DDMM is either done by striking the drums in time or by making dance steps respectively. For this purpose, an acoustic incentive is provided in the form of a metronome. The timings of all user actions are then compared with those of the metronome beat and corresponding notes of a melody are played if they match. At the same time, the position of the arms is evaluated by the camera system and passed as a further parameter with which the key of the piece can be changed. Thus, for example, children can explore the difference between major and minor keys by placing their arms in six different positions, which are based on the upper part of the circle of fifths. After all this first prototype contained the basic functions and characteristics required for the project, but an evaluation with children had not yet been done. Also, the planned visual feedback, which should complement the concept, was not implemented at that time.

1.3 Enhancements

Several improvements were implemented after the completion of the first prototype. In particular, the additional visualization should be mentioned here. Instead of using abstract graphics we followed the concept of interactive demonstration to realize this visualization, since it should motivate the user and at the same time give guidance. Therefore, we created a new plugin with virtual companions, who will teach the children what to do. Since we want to find out, if the children would rather use a realistic than a non-realistic form of representation, we furthermore implemented two versions of this new plugin (see figure 1) in order to compare them.

![Figure 1: additional visualisation modules (left: realistic version, right: cartoon version)](image)

The application is controlled by plugin “Music” which communicates through a specially defined protocol and so it is possible to let the characters drum and dance in strict time. In addition, the colors of the visualization adjust to the transmitted keys of the melody. As a result, the sky gets darker, when the children switch the tune from major to minor and vice versa. (Steinmeier & Becking, 2017).

2. Related Work

There are also some approaches, such as in the work of Sarasua, which are reminiscent of DDMM’s approach and pursue a similar goal. (Sarasua, 2013) However, the overall focus of these projects is usually not on musical education in the preschool age, but rather on older children or people with impairments. Nonetheless, it can be said that there are significant research activities that deal with the investigation of novel instruments and concepts designed to motivate children to practice music making.
The work of Myriam Desainte-Catherine et al. for example belongs to a series of projects, which assume that it is particularly motivating to take up aspects of computer games and connect them with musical activity. (Desainte-Catherine, Kurtag, Marchand, Semal, & Hanna, 2004) Varni et al. pursue a similar concept, but use on a complete computer game and not just a partial aspect. (Varni, Volpe, Sagoleo, Mancini, & Lepri, 2013) In this game, the children are told a story in which they can interact to help a magician fill the world with sounds that have been lost. The findings of Blaine and Perkis are also interesting for the development of DDMM since they also chose drumming interfaces as the simplest option for collaborative improvised music making. (Blaine & Perkis, 2000) In their study, up to six users are able to play an instrument consisting of several electronic drum pads that are connected to the computer by an Alesis Trigger module. In addition to this, they also developed and tested approaches for a visualization, which should provide feedback and motivation. Their concept does not, however, allow each musician to play an individual instrument and instead create a global mixture from their inputs. In this regard Blaine and Perkis argue that some of their subjects had problems identifying their own contribution to the composition. Also the system “WamBam” (Jense & Leeuw, 2015) should be mentioned in this context. In their paper, Jense and Leeuw describe the results of several prototypical development stages: An electronic instrument, which is similar to a drum and designed for the music-therapeutical treatment of patients with severe mental disabilities. Their research confirms, that the idea of making music with the interface of a drum is useful and activating and their results suggest that this aspect will be well received by the children, since it supports their auturgy and perception despite their limitations.

3. Evaluation

In principle, the evaluation of DDMM was divided into two phases. The initial state of the system (the previous prototype) was investigated in a first test run. Based on the observations from this test run and the concept for the further development we implemented the necessary improvements (e.g. visual feedback), which was then evaluated in the second test run. In both states the system was evaluated with children from three different age groups: at first with participants of a voluntary drum group, which is offered weekly at a local elementary school within the framework of the open all-day school (OAS), then in a music-related parent-child group and the last test took place in a children's day-care center. In order to discover, how children deal with DDMM, to what extent they interact self-determined and whether they have (subjectively speaking) fun, we used the method of open participating observation as the main component of this evaluation. Furthermore, we combined this procedure with other methods in order to reveal further connections. On the one hand, we asked the parents / teachers using a standardized form in order to obtain information on the age and behavior of the children during their everyday life. On the other hand, the children were asked about their opinion on the system, which we did in the form of a group interview after trying and exploring the system. The test phase itself also took place in this group, whereby a maximum group size of 15 persons was selected, since it corresponds roughly to the day care group size. The children knew each other and were able to interact with their usual friends in smaller test groups, while the other children of the group watched or joined in. The tests and interviews were documented with the help of video recordings and afterwards both children's utterances and their behavior were transcribed in an anonymous form, in order to examine them in various aspects. In addition, during each test, the system itself measured the success of the children playing.

3.1 First Test Run

In general, the structure and process of evaluation is similar in all facilities so that the comparability of the results is given. However, there are minor differences in available space or group strength. We arranged the components of the system in the room so that observation is possible from several sides. After the technical setup, the further course is then divided into three phases (introductory phase, test phase and interview phase). It could be observed, that most children reacted to DDMM with great joy and very enthusiastically and they urged to test the system right from the start. In only a few exceptional cases the children did not dare to try DDMM themselves and preferred to look. It was also observed in
all groups how attention and concentration were lost towards the end of the test phase. Thus, the first test runs were much more careful and explorative than the last ones. Furthermore, it could be observed, that the children usually played in strict time by chance rather than coincidentally and it appeared, that they play the drums uncontrolled. In many cases, they did not hit the beat over several bars. Among the dancers, it was often observed that the children danced rather freely (breakdance, couple dance). Altogether, only a small part of the children tried to produce music by stomping and only a few tried to change the tune with their arms. At the beginning, it was basically the case in all facilities, that most of the children were drumming and only a few wanted to dance. This changed in the course of the evaluation and towards the end, more children wanted to dance than do the drumming. In the following interview the children in all the tested facilities agreed, that they would describe the machine as "goooood" or "niiiiiice". Regarding the question if whether drumming or dancing was easier, there was a clear tendency towards drumming, but in each group at least one child thought dancing was easier.

A total of 19 girls and 16 boys from 3 to 8 years of age took part in this test, with more than half of the children in preschool age. Further analysis of the data from the questionnaire shows that all children at least show interest in making music regarding individual aspects. According to the data of educators and parents more than 80% of the children sing songs on occasion and almost 90% play on an instrument. The comparison of subjective observations with questionnaire data also shows a similarity: the children (mostly girls) who occasionally dance in their everyday life often also raise their hand for the role of the dancer during the test of DDMM. In order to evaluate the log files generated during our tests, an analysis program was coded, which maps the information given by the text based log files (e.g. child has hit in strict time / not hit in strict time or not hit at all) as numerical values and writes them into a table. Figure 2 shows the graphical representation of the data from one test group during both iterations:

![Graphical representation of measured data from one exemplary test group.](image)

This and the other diagrams often show a similar picture: the curve of the dancers (Kinect Musician) drops almost constant in most cases, while the drummers’ curves (Drums Musician) often take course around the zero baseline. The few exceptional points of the curve show the moments when the children were stomping and temporarily matching the metronome. These objective values also coincide with the subjective observations. In many groups the children, as already described, made quite different dance movements and accordingly the results of the measurements were to be expected. The same applies to the frequency of the key changes provoked by the different arm positions. Based on the observations we made, it should also be noted that the kindergarten children were already familiar with this practice in a didactic-methodical learning environment. Thus, their participation was more calm and concentrated than during the tests with the children from OAS and parent-child group. This allows the conclusion that certain groups need a clear setting for working with DDMM, such as, for example, a manageable group size adapted to the age structure, to prevent longer waiting times for this activity, avoid unrest and increase the concentration. Likewise, an intensive introduction such as trying out drumming, dancing and stomping movements, elaboration of rules for social behavior (e.g. the passing on of the drums) and the motivation for self-control of the child by the pedagogues can be helpful.

### 3.2 Second Test Run

In principle, both a similar structure as well as a nearly identical procedure is used in the second test iteration. In this test run, the variances regarding the structure are mainly due to the additional placement of a screen, which is used to display the visualization application. Based on the impressions of the previous test run, there are also some minor changes for the process. Since it was noticed that some of
the functions and properties of TTMM could not be perceived by the very free approach in some cases, more information and suggestions regarding the handling of the system should be given during this test iteration. As for the visualization, at first the cartoon version is used and then replaced by the realistic designed scene after a few test groups. In the last passages, the children are finally allowed to dictate which version should be used.

On the one hand, it could be observed that by means of the structured procedure with advices and instructions the children could be better informed about their actions. On the other hand, the dance behavior of the children also changed due to the visualization. It was observed that there was less free dancing and much more stomping, since the children tried to copy the behavior of their visual guides. Due to the color changes of the visualization, the children were also strongly encouraged to try different arm positions and analyse the resulting changes. Like in the first test iteration, at the beginning none of the children wanted to dance. However, this changed after the first group. Since some of the children from the kindergarten group already knew DDMM from the first iteration, during the subsequent interview phase it was possible to ask whether they think the system is either better with the new visualisation or if they would not need it. Everyone agreed that making music with the help of the additional visual guidance and feedback was better and easier.

A total of 17 children from four to seven years of age took part in this second test run and, as in the first run, more than half of the children participating were in the preschool age. In the first review of the evaluation results of the questionnaires it can be seen, that the diagrams of both institutions differ significantly from each other this time. Regarding the behavior and the measured success of the dancers, a significant improvement could be achieved instead. While the children in the first test run rarely managed to stomp in time, the graphical evaluation of the second test run shows much more success. In two cases, the curve of the dancer even takes course above the two drummers’ curves (see figure 2). Looking further at the data, the average number of changes in the position of the arms shows that with the use of the visualization not only the frequency of the stomping movements but also the amount of arm movements increased. On average, children change the tune of the composition three times more often than during the first test iteration. Both of these changes can also be seen in the measurements. In contrast, however, no changes can be measured with respect to the drummers: their curves are very similar to those of the first test run. From this it can be deduced that the process of drumming was self-explanatory to them and further aid would not have been necessary.

4. Overall Results and Conclusion

The additional optical feedback was intended to motivate the children to an even greater extend. In the two test runs, both development stages were therefore investigated and the most important results and observations were presented in this paper. We were able to show that the general approach of motivating children to create music together through the use of a novel system is well received and can lead to the desired success. Similar results were obtained from the evaluation of the further developed version, whereby it could also be measured and observed that the visual feedback motivated children to dance and explore DDMM even more. Unlike we feared, the occasional lack of acoustic feedback - when the children do not play in strict time, meaning they hit the drum but produce no sound - does not seem to lead to great confusion and frustration. With regard to the question of research we wanted to answer, it can thus be concluded that the system appears to be suitable for the use in early music education. However, the extent to which the application has a learning effect that is actually usable should be examined in further, more extensive investigations. A further result of the two test runs are the opinions of the children on the question of what else could be added to the system. Furthermore, from the observations some other possibilities of improvement can be derived, such as an additional free dance mode and a further development of the drum.

5. Future Work

The feedback from the institutions as well as the results of the evaluation show that it might be useful to carry out further test runs and possibly also to analyse the already collected data regarding other aspects.
Thus, further questions, such as the achievable learning effect or the suitability of the system for even younger children, could be investigated. Since the results of the previous tests have shown, among other things, that a more structured approach can lead to better results, a detailed pedagogical concept should be developed beforehand. Additionally further enhancements, based on the observations we made, with a subsequent evaluation is planned for the future. For example, recently an additional DDMM plugin was developed by other students, which offers the possibility to query different areas of a drum. Thus it would be feasible that drumming children could have a greater influence on the design of the common piece of music. This also enables the important educational goal of self-determination and self-efficacy of the child. (Schwarzer & Jerusalem, 2005) In addition, the project is supposed to lead to further research projects, particularly in the field of curative education. However, this is in principle only one of many possible applications in which the system we developed can be used.

References


Effects of Peer Interaction on Web-Based Computer Programming Learning

Nuttaphat ARUNOPRAYOCH\textsuperscript{a}, Chih-Hung LAI\textsuperscript{a}, Pham-Duc THO\textsuperscript{a} & Jing-San LIANG\textsuperscript{b}

\textsuperscript{a}Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien, Taiwan, Republic of China
\textsuperscript{b}Taipei Municipal Sanxing Elementary School, Taipei, Taiwan, Republic of China
*nuttaphat.aruno@gmail.com

Abstract: This paper aimed to investigate an effectiveness of peer-interaction strategies on web-based computer programming. Totally, thirty-three computer science students participated in the programming course. The learning achievement was compared by using an e-learning system and a peer-interaction system. Moreover, the correlations between peer-interaction functions and their students’ learning achievement, and the students’ attitudes toward the new system were examined. The results demonstrated that there is no significant difference between using traditional e-learning system and the peer-interaction system. Furthermore, peer-interaction functions had a positive relationship with their learning achievement. Ultimately, it is found that students had positive attitudes toward the peer-interaction system.

Keywords: peer-interaction, peer assessment, computer programming, online learning system

1. Introduction

Computer programming has become a must-learnt subject since the world has already entered digital era. Thus, Introduction to Computer Science is a required fundamental course that students must take as a first course (Flores, Barron-Cedeno, Moreno, & Rosso, 2015). Learning computer programming language is not an easy task because of its challenges, for example, problem-solving ability, the idea of mathematics, and computer proficiency (Flores et al., 2015).

To make learning computer programming more interesting and efficient, a considerable amount of research has been conducted by using different strategies such as peer-assisted learning, question generating, source code reusing, web-based learning, unidirectional and reciprocal teaching, and online annotations (Altintas, Gunes, & Sayan, 2016; Flores et al., 2015; Kalelioglu, 2015; Lluxton-Reilly, Denny, Plimmer, & Bertinshaw, 2011; Shadiev et al., 2014; K. Topping, 1998). Some research clarified that learning with peers plays an important role in today’s education to improve students’ understanding of the central concepts of a subject. Moreover, it helps promote social skills which are needed in society; therefore, such strategy shall be applied to computer programming learning (Altintas et al., 2016; K. J. Topping, 1996; Yeh, Hung, & Chiang, 2017). The results from the previous research showed that students had an outstanding academic performance by using the strategies mentioned above (Altintas et al., 2016; Flores et al., 2015; K. Topping, 1998; K. J. Topping, 1996; Yeung & Nguyen-Hoang, 2016).

Interestingly, there is one strategy that stands out of others; it is called peer-interaction strategies which allow students to interact with their peers to complete given tasks. In this study, the researchers proposed the use of Peer-interaction Programming Learning System (PIPLS) to support computer science students in learning computer programming language. The study investigated three research questions listed as follow:

- Does the use of PIPLS improve students’ academic performance?
- Are there any correlations between provided peer comments and learning performance?
- Are there any correlations between quality of received peer votes and learning achievement?
- What are students’ attitudes toward the peer interaction learning system?
2. Literature Review

Peer interaction is significant in a class (Yu & Wu, 2012). Yu and Wu (2012) found that learners who were using a system based on peer-interaction context tend to use both learning strategies and deep learning approaches; students chose different techniques applied to each given task in order to have it successfully answered (Russell, Van Horne, Ward, Bettis, & Gikonyo, 2017). Indeed, there are prior studies have provided evidence supporting the claim that involving peers in learning is essential (Altintas et al., 2016; Flores et al., 2015; K. Topping, 1998). Topping (1996 and 1998) examined the effectiveness of using peer tutoring and assessment in higher education; he found that these methods promoted students’ understanding and academic achievement; therefore, further research should focus on these approaches accompanied with using future technology.

In 2011, StudySieve was introduced by a group of researchers at The University of Auckland to evaluate higher-order thinking and their perception toward StudySieve (Luxton-Reilly et al., 2011), it is a system that allows students to generate multiple-choice questions. Its capabilities are intriguing seeing that the system can provide an environment allowing students to answer and comment others’ questions and answers. The results showed that students were active to learn contents deeply by creating questions and answering others’ questions during the system-used courses.

However, its limitation, which is clearly seen, was that besides multiple-choice questions; other types of questions such as essay, fill-in-the-blank, and especially coding with auto-grade function were not included. To tackle those limitations, a system that has a variety of questions types must be developed. Consequently, PIPLS has been developed by Tho Pham-Duc and his team under the guidance of Prof. Lai Chih-Hung (Lai & Tho, 2016).

3. Methods

3.1 Research Design

This study was carried out in a simple design by using quantitative approaches. It was conducted in pre-test/post-test comparison between Test 1 (E-learning) and Test 2 (PIPLS). The independent variable of this study was a group of participants. The dependent variables of the study were the test scores from two different teaching methods: traditional e-learning and PIPLS. The participants’ perceptions toward the system were examined during a final test phase as shown in Figure 1.

![Figure 1. Research framework](image)

3.2 Participants

A sample of thirty-three undergraduate students of Computer Science and Information Engineering who enrolled in Introduction to Computer Programming II were selected; in other words, every student in the class was asked to participate. There were 25 males and 8 females. All students were approximately aged 20 years old.
3.3 Learning Contents

The class, Introduction to Computer Programming II, was taught separately as a 3-hour lecture and a 3-hour laboratory per week, using an e-learning system and PIPLS as mediators:

- During the use of e-learning, for the 3-hour lecture, the content was introduced with PowerPoint to explain the main concept of each lesson; in the 3-hour laboratory class, students were asked to use a compiler to complete given tasks, then submit a code file on e-learning system.
- During the use of PIPLS, even if PowerPoint was still used to introduce the main concept to students, the 3-hour laboratory was conducted with PIPLS allowing students to interact with their peers inside and outside the classroom to resolve assignments.

3.3.1 Peer-Interaction Programming Learning System (PIPLS)

It is a web-based learning system based on the open source Question2Answer system (Lai & Tho, 2016) modified with further development and in-depth web programming customization which allows an instructor to create a class on the system, PIPLS’s interface as seen in Figure 2. It has been designed; the user interface, to provide question generating and peer interaction functions such as:

- Question generation types: essay, multiple-choice, short answer, true-false, fill-in-the-blank, and coding with both an automatic and a semi-automatic assessment from staff or peers.
- Peer-interaction functions: peer assessment (comment) allows the student to give sophisticated suggestions to peers’ questions and answers, chat room, asking for help (letting peers know that help is needed), voting function (up-vote and down-vote) which allows students to assess the quality of questions and answers, and providing open comment.
- Others: questionnaire, leaderboard, and achievement trophies.

![Figure 2. The PIPLS’s main page](image)

3.4 Experimental Procedures

This study was carried out in three phases. Students were assigned to take a computer programming proficiency test at the end of each phase. During phase 1 students received a traditional lecture by using PowerPoint as a teaching material, and used an e-learning system to submit assignments. On the other hand, in phase 2 and 3 students were still taught by a traditional method, but were asked to use PIPLS as a tool to complete assignments by using peer-interaction functions provided. The first two phases continued for a period of five weeks, and the last phase continued for a period of eight weeks; therefore, this study had been conducted for 18 weeks consecutively started from the beginning of the 2016 spring semester. The questionnaire concerning perceptions was digitally provided, along with the final test, at the end of the 18-week experimental period.

3.5 Data Collection and Instrument

In this study, PIPLS was used as a data collection tool in order to bring answers to research questions. Creating an account to be used as an identification accessing to the system was essential. An account
contains a student’s information. Furthermore, PIPLS collects every data from an account in the form of a database powered by MySQL, the frequencies of functions usage, the time spent on the system, answered questions, call for help, comments, voting, etc. Thus, the collected data derived from the database were statistically analysed.

The students’ attitudes toward the system were measured by using 5-Likert scales (1 = strongly disagree and 5 = strongly agree) questionnaire provided in the system. The questionnaire, based on Kooresse’s research (Kooresse, 2012) according to the information appeared in programming assistance tools (Luxton-Reilly et al., 2011), consisted 16 questions regarding the students’ preference between an e-learning and PIPLS.

3.6 Data Analysis

To reveal whether there was an academic difference between pre-test and post-test or not, paired-samples t test was statistically calculated with the data collected from PIPLS’s database using SPSS. To test whether there was a relationship between the use of peer-interaction functions (e.g., comment and voting) and final scores, Pearson’s correlation and simple regression were calculated. In addition, the students’ attitude toward PIPLS were explained by descriptive statistics. It should be noted that 5% had been determined as the significance level.

4. Results and Discussion

To investigate the effects of PIPLS on exams, paired-samples t test was applied. Thus, the null hypothesis which there is no statistically significant difference between the means of Test 1 (e-learning) and Test 2 (PIPLS) regarding their academic exam scores was tested.

As seen from Table 1, there is no statistically significant difference between the means of Test 1 (M = 80.95, SD = 29.71) and Test 2 (M = 83.68, SD = 18.22) regarding exam scores at the .05 level of significance (t = -.44, df = 32, n = 33, p = .66). However, on average students’ scores after using PIPLS were higher than their previous scores according to the difference of means. It should be noted that there was a violation found in paired-samples correlation.

According to t-test results, learning programming via PIPLS did not cause any differences from using e-learning regarding test scores. There, only, were slight increases in the means of Test 1 and Test 2’s scores. Even if PIPLS’s mean was higher than e-learning’s mean, the occurred violation should be noted. More importantly, the results may be caused by the number of participants which were only 33 students involved and the different difficulties of the concerned exams. Similarly, the results from the research using peer-assisted learning showed that there was no statistically significant difference between the means of the experimental group applied peer-assisted learning and the control group applied traditional teaching methods regarding their academic scores (Altintas et al., 2016).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n</th>
<th>df</th>
<th>Test 1 (e-learning)</th>
<th>Test 2 (PIPLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>33</td>
<td>32</td>
<td>80.95, 29.71</td>
<td>83.68, 18.22</td>
</tr>
</tbody>
</table>

To answer the questions that whether the frequency use of peer-interaction functions (e.g., comment and voting) improve students’ final score or not, correlation and regression analysis was statistically performed.

Table 2 demonstrates relationships between peer-interaction functions and the final scores. The results showed that students, who had a high amount of comments toward peers’ questions and answers, successfully achieved higher final scores compared to those who did not (r = .35, df = 31, p = .049). It was also found that students whose answers were upvoted, which means received good feedback for the most informatively-useful answers, by their peers performed significantly better than those who did not receive (r = .36, df = 31, p = .039).
The results in this experiment showed that students benefit from peer interactions in terms of learning. Even though there was no significant difference regarding their final scores, it can clearly be noticeable that peer-interaction functions in PIPLS (e.g., comment and voting) had positive relationships.

Furthermore, the predictive powers of the frequency use of peer-interaction functions on the students’ final scores by using simple regression analysis were presented in Table 3. The results showed that giving peers comments was a significant predictor of final scores, $\beta = .35$, $t(31) = 2.05$, $p = .049$, $R^2 = .12$. Moreover, it demonstrated that receiving up-vote was a significant predictor of final scores, $\beta = .36$, $t(31) = 2.15$, $p = .039$, $R^2 = .13$. It can be said that giving peers comments and receiving up-votes could be a factor that affects students’ academic achievements regarding their final scores in this case.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Final score</th>
<th>Comment</th>
<th>Receiving up-vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>.35, .36</td>
<td>.12, .13</td>
<td>.049*, .039*</td>
</tr>
<tr>
<td>Final score</td>
<td>33, 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.12, .13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>.049*, .039*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mentioned results are similar to prior research, for example: peer instruction (PI) applied to calculus and algebra-based introductory physic course indicated that students mastered in both conceptual reasoning and problem-solving skills upon the implementation (Crouch & Mazur, 2001), and involving peers in learning process as reciprocal teaching indicated a significant result of better improvements (Shadiev et al., 2014; Yeh et al., 2017).

Another finding of this study was that students had good perceptions and preferences toward using PIPLS compared to using e-learning ($N = 33$, $M = 3.95$), it should be noted that 5-Likert scales were used; 0 indicates ‘strong disagree’ and 5 indicates ‘strongly agree’. The questions were made to explore the students’ thoughts toward PIPLS and their motivation for future use. The examples from the questions as follow:

- The function of viewing others comments helps me learn better.
- The function of viewing others comments motivated me to use the system.

According to the results of attitude toward PIPLS, most students agreed that using PIPLS was useful for learning programming compared to an e-learning system described by interpreting $M = 3.95$ which was statistically close to 4 indicated ‘agree’. Students mentioned that explaining and providing suggestions to peers could help them to learn program concept deeper and, to find some mistakes in their codes.

5. Conclusion

The main purpose of this study was to explore the effects of peer interaction on web-based programming language learning system. The main conclusions of this study are summarised as follows. Firstly, the slight improvement was found after students used PIPLS in terms of their second test scores, however statistically there was no significant difference between two sets of scores. Secondly, frequently using peer-interaction functions provided by PIPLS did affect students’ final scores. Lastly, most of the students perceived that peer-interaction methods supported by PIPLS were more useful for learning computer programming languages compared to an ordinary e-learning.
To be informed, this study is primarily limited by its sample size and experimental methodologies; therefore, a larger sample size would have benefited the results of the study to conduct the comparison of a control group and an experimental group, and the implementation could have been longer. Other teaching methods would have been more useful than conducting only traditional methods (i.e., using Powerpoint’s slides and teacher-centered) to deliver knowledge (Matthiassdöttir, 2006). Moreover, the differences between genders: male and female, and the time spent on the system might be critical factors as well.

In a nutshell, peer interaction helps students not only to promote their understanding of programming concepts but also improve their academic achievements. Thus, the contribution of such methods based on web-based programming learning along with engaging students to frequently use peer-interaction functions should be further investigated for future research.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology of the Republic of China under contact number: MOST 105-2511-S-259-004-MY3.

References


233
Exploring Lag Times in a Pair Tracing and Debugging Eye-Tracking Experiment

Maureen VILLAMOR* & Ma. Mercedes RODRIGO*

*Ateneo de Manila University, Philippines
bUniversity of Southeastern Philippines, Davao City, Philippines
*maui@usep.edu.ph, mrodrigo@ateneo.edu

Abstract: This paper investigated the leader/follower patterns that possibly occurred in a remote pair programming eye-tracking experiment. We intended to build the profile of the initiator and the follower and explore the lag times inherent to pairs categorized based on prior knowledge using the diagonal recurrence profile. Findings revealed that in a pair programming setup, the initiator was the low prior knowledge participant. We defined the “initiator” as the one who encountered a problem in the code first and hence initiated the contact to ask for help, and the “follower” was the one who responded to help. The characteristic lag times based on prior knowledge were 2.33 seconds for both high prior knowledge pairs, 1.96 seconds for both low prior knowledge pairs, and 1.51 seconds for mixed prior knowledge pairs.

Keywords: Leader/follower, lags, eye-tracking, pair programming, diagonal recurrence profile

1. Introduction

Schilbach (2015) defined joint attention as “attending to something together with someone and being aware that ‘we both’ are attending.” It establishes a form of cooperative behavior that relies on shared intentions (Pfeiffer et al., 2014) and denotes following the direction of another person’s gaze (Bayliss, Pellegrino and Tipper, 2005). For joint attention to occur there must be an initiator directing someone else’s gaze toward a particular target and a follower who will go along with the gaze cue provided.

Cross-recurrence quantification analysis (Marwan and Kurths, 2002) is used to quantify how frequently two systems exhibit similar patterns of change in time. It produces a cross-recurrence plot (CRP), which has been used to analyze the coordination of gaze patterns between individuals and determine how closely two collaborators’ gaze follow each other. One important measure that can be derived from the CRP is the diagonal recurrence profile (DiagProf), which can be interpreted as a lag profile that reflects co-occurrence patterns between utterances of varying relative lags (Warlaumont et al., 2010). This measure was used in the analysis of linguistic and eye-movement coordination in the studies conducted by (Richardson and Dale, 2005; Richardson, Dale and Kirkham, 2007).

This paper used the diagonal recurrence profile to answer the following research questions: (1) What is the profile of an initiator and a follower in a pair tracing and debugging eye-tracking experiment? and (2) What is the characteristic lag time between pairs of participants categorized according to prior knowledge? In a prior study we conducted (Villamor et al., 2017), we speculated that the mixed-prior knowledge pairs may have a presence of leader/follower patterns where we hypothesized that the high prior knowledge participant was the one taking the lead in debugging and telling the low prior knowledge participant what to do. We wanted to validate in this study if this hypothesis is true.

2. Methods

The study was conducted in four private universities in the Philippines. Thirty two (32) pairs of participants aged 18-23 years old who were in their 2nd year to 4th year level in college and had taken the college-level fundamental programming course were recruited to participate in this study. For a detailed explanation on the structure of the study, see Villamor and Rodrigo (2017).
A diagonal recurrence profile was constructed for each pair for every program by implementing the function `drpdformats` from the crqa package of Coco and Dale (2013) for R. Each sampling unit is 33 milliseconds. For this data, the radius was set to 5% of the maximal phase space diameter (Schinkel, Dimigen and Marwan, 2008). More information about the diagonal recurrence profile can be found in Fusaroli, Konvalinka and Wallot (2014).

3. Results and Discussion

To determine the initiator and the follower between the participants with a low- and high-prior knowledge, only the mixed prior knowledge pairs (LH) were considered. For consistency, the pairs were ordered starting with the low prior knowledge (LPK) participant followed by the high prior knowledge (HPK) participant. The function `drpdfromts` was used to compute the diagonal recurrence profile of the two fixations sequences per program for each pair under different categories. This function returns a recurrence profile with the length equal to the number of lags considered, the maximal recurrence observed between the two fixation sequences, and the lag at which it occurred. However, for this study we focused only on the lag at which the maximum recurrence was observed.

Results revealed that the initiator and follower were the LPK participant and the HPK participant, respectively. The LPK participant’s gaze was on the average ahead of the HPK participant’s gaze by 1.51 seconds 53.16% of the time. The baseline delay is 1.89 seconds. This lag time is still within the range reported in Richardson and Dale (2005) where the follower needs a lag of two (2) seconds to be maximally aligned with the speaker’s eye movements.

One possible interpretation for this finding is that in programming tasks, it is usually the low performing students who ask help from their more experienced peers. This is a common scenario observable in class laboratory exercises. The gaze direction in this study was initiated by the LPK participant, possibly in an attempt to seek help from the HPK participant. It could be that in most cases the LPK participant encountered problems in the code and pointed them out to the HPK participant who followed the direction of the LPK participant’s gaze. Hence, in this case, the “initiator” was the participant who encountered a problem in the code first and instigated contact with the other to ask for help, and the “follower” was the one who responded to help and followed the gaze of the initiator.

Of the 31 pairs (one pair was discarded), 11 had both high prior knowledge, 6 had both low prior knowledge, and 14 were mixed pairs. We will refer to these categories as HH, LL, and LH. The lag time at which the maximum recurrence occurred for every program based on prior knowledge were averaged. The aggregated results were examined to find differences among the categories, which entailed looking at incidences of long and short lags. A lag was long if it was equal to or greater than the mean plus one standard deviation; and short, otherwise.

The HH, LL, and LH pairs exhibited coupled gaze patterns lagged at about 2.33 seconds, 1.96 seconds, and 1.51 seconds, respectively. Further results showed that HH pairs had average to long lags, the LL pairs had short to average lags, and LH pairs had a variation of lags but majority of the lags were below the mean. HH pairs had the longest lag while the LH pairs had the shortest.

Results showed that the HH pairs took longer for their gazes to focus on the same target, whereas it took faster for the LH pairs for gaze coupling to take place. One possible explanation for the longest lag for HH pairs is that the HH pairs may not have collaborated as much compared to other relationships. Findings from our previous study confirmed this speculation. HH pairs had the lowest average recurrence rate (Marwan and Kurths, 2002). It could be that the HH pairs did not feel the need to collaborate more often because they were already confident with their work. It is also possible that they were engaging in more divergent episodes (Sharma et al., 2012). For the LL pairs, two persons who are both inexperienced in debugging and who are struggling with program comprehension frequently may have difficulties trying to understand and locate in the code what the other one is trying to point out. Students who are low performers or have low prior knowledge do not even know correct programming terminologies, which is probably one of the reasons why the LL pairs took longer for their gazes to overlap. Lastly for the LH pairs, since it was the LPK participant who initiated the gaze direction, as revealed in the previous section, it was easy for the HPK participant to follow suit since the HPK participant has more experience in debugging and thus can easily understand and find in the code what the LPK participant was referring to.
4. Summary and Conclusion

The goal of this paper was to build the initiator/follower profiles between pairs of novice programmers and investigate the lags between pairs categorized according to prior knowledge to find out the delay that maximally aligns their gazes. Results showed that the initiator majority of the time was the low prior knowledge participant. This result was counterintuitive to our hypothesis from our previous study. It turned out that the gaze initiation was possibly a result of the low prior knowledge participant asking for help from the high prior knowledge participant. The lag times of the pairs categorized according to prior knowledge varied, which was within the range of the lags recorded in the study conducted by Richardson and Dale (2005).

These preliminary findings are significant because it paves the way for us to determine if the similarity between participants’ gazes can result in similar cognitive states particularly in programming. The lags can inform us how students with different prior knowledge interact and collaborate through their gazes. Future work will look into the impact of the gaze coupling with respect to how the follower understood the initiator to reflect the follower’s attentiveness and the success of their communication by exploring both the eye-movement and discourse data.

References


The Research of Interaction Performance of Intercultural Communication in Computer-supported Collaborative Learning

Yi WEI*, Li-Jie WU & Yi-Ling HU

Department of Educational Information Technology, East China Normal University, China
*toweiyi@163.com

Abstract: This study mainly investigated the relationship between the interaction performance and intercultural communication competence (ICC) in computer-supported collaborative learning (CSCL). We designed two instruction activities which can support the students collaborated with each other in a cross-culture environment. And we used teacher assessment and student questionnaire to evaluate the students’ interaction performance and ICC. Correlation analysis and independent sample test were performed to process the collected data. The findings revealed Dissonance and Negotiation are the two phases during the interaction process correlated to many factors of students’ ICC. And the interaction performance in CSCL has no significant difference between native speakers and non-native speakers.

Keywords: CSCL, interaction performance, ICC, intercultural sensitivity, intercultural effectiveness

1. Introduction

With the development of technology and the continuous progress of globalization, intercultural communication activities have become increasingly frequent and it can take place at anytime and anywhere. Computer-supported collaborative learning provided a convenient way for intercultural communication among students from different countries (Portalla and Chen, 2010). But the magnitude of these communication challenges increases as cultural differences among communicators widen. It requires students improve their ICC constantly to adapt to the development of this society.

Computer plays an important role in collaboration, some designs for CSCL include situations both working at a distance and face-to-face, as well as mixtures of synchronous and asynchronous collaboration (Goodyear, Jones and Thompson, 2014). ICC is a set of abilities (such as the knowledge, motivation, attitudes, and skills, etc.) to interact effectively and appropriately with members of different cultures (Peng and Wu, 2016; Spitzberg and Changnon, 2009; Wiseman and Koester, 1993).

In this study, there are two research questions: (1) Is there a correlation between the interaction process in CSCL and the students’ ICC? (2) Is there any significant difference of the interaction performance between native speakers and non-native speakers?

2. Method

15 international high school students participated in this study (males=9, females=6). They came from 6 countries (American=1, Canadian=1, Chinese=7, Korean=2, New Zealander=1, Singaporeans=3). The students were divided into three groups and each group had five students of different levels. We designed two activities, and the scheme was “Professional Term”. In the first activity, the students discussed how to explain these difficult terms more clearly. And in the second activity, students collaborated with their group members to make a PowerPoint to show their results. They cooperated to complete the tasks with computers, and the teacher gave some guidance to each group. All of them can speak English in the activities. The teacher observed and scored the students’ interaction performance during the activities. And the students filled in the questionnaire at the end to measure their ICC. The
interaction process includes five phases: Sharing/Comparing, Dissonance, Negotiation/Co-construction, Testing Tentative Constructions, and Statement/Application of Newly-Constructed Knowledge (Gunawardena, Lowe and Anderson, 1997). The full score of each phase was 5, a score of 2.5 or more than 2.5 was good. The questionnaire includes two sections, the first part collects respondent information such as grade, class, gender and country, and the second part involves 44 items about intercultural sensitivity (24 items, including interaction engagement-IG, respect for cultural differences-RD, interaction confidence-IC, interaction enjoyment-IJ, and interaction attentiveness-IA) and intercultural effectiveness (20 items, including intercultural effectiveness-IE, behavioral flexibility-BF, interaction relaxation-IL, interactant respect-IR, message skills-MK, identity maintenance-IM, and interaction management-IN) (Chen and Starosta, 2000; Portalla and Chen, 2010). A 5 point Likert type scale (1 for strongly disagree and 5 for strongly agree) was used to rank the level of disagreement and agreement. And we processed these data by SPSS 19.0.

3. Result and Discussion

3.1 The Correlation between Interaction Performance and ICC

We used P1, P2, P3, P4, P5 to represent the five phases of the activity. The correlation between interaction performance of each phases and ICC is shown in Table 1. The results show that P1 was positively correlated with IR, but no significant correlation with other factors of ICC was found. P2 was positively related to RD and IJ, and it was significantly related to IR and IM. P3 correlated significantly with IR, and it also positively related to IL, MK and IM. Although P4 had a significant positive relationship with IR, it wasn’t related to other factors. P5 was related to IL, and it also significantly correlated with IR. So, the second (Dissonance) and the third (Negotiation) phase in the process are worthy of more attention. In these two phases, the instruction design should be more clearly and detailed, and the teachers should give more appropriate scaffolding and guidance to student to help them accomplish these tasks successfully. In addition, IR was related to each phases of the activities. IR had a relatively strong correlation with P1, and it was significantly correlated with the other four phases. When we train students’ ICC, we should focus on the cultivation of interactant respect ability.

Table 1: Correlations among variables of interaction phases and ICC.

<table>
<thead>
<tr>
<th>Variables</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG</td>
<td>.381</td>
<td>.486</td>
<td>.504</td>
<td>.172</td>
<td>.247</td>
<td>3.53</td>
<td>0.364</td>
</tr>
<tr>
<td>RD</td>
<td>.299</td>
<td>.595*</td>
<td>.345</td>
<td>.121</td>
<td>.205</td>
<td>3.96</td>
<td>0.602</td>
</tr>
<tr>
<td>IC</td>
<td>.304</td>
<td>.477</td>
<td>.474</td>
<td>.228</td>
<td>.309</td>
<td>3.73</td>
<td>0.554</td>
</tr>
<tr>
<td>IJ</td>
<td>.243</td>
<td>.531*</td>
<td>.400</td>
<td>.133</td>
<td>.239</td>
<td>3.8</td>
<td>1.06</td>
</tr>
<tr>
<td>IA</td>
<td>.044</td>
<td>-.048</td>
<td>.111</td>
<td>.251</td>
<td>.031</td>
<td>3.6</td>
<td>0.726</td>
</tr>
<tr>
<td>BF</td>
<td>.045</td>
<td>.122</td>
<td>-.005</td>
<td>.002</td>
<td>.269</td>
<td>3.23</td>
<td>0.522</td>
</tr>
<tr>
<td>IL</td>
<td>.402</td>
<td>.461</td>
<td>.609*</td>
<td>.457</td>
<td>.529*</td>
<td>3.63</td>
<td>0.609</td>
</tr>
<tr>
<td>IR</td>
<td>.581*</td>
<td>.776**</td>
<td>.765**</td>
<td>.773**</td>
<td>.724**</td>
<td>3.87</td>
<td>0.602</td>
</tr>
<tr>
<td>MK</td>
<td>.297</td>
<td>.393</td>
<td>.528*</td>
<td>.316</td>
<td>.270</td>
<td>3.05</td>
<td>0.326</td>
</tr>
<tr>
<td>IM</td>
<td>.471</td>
<td>.741**</td>
<td>.563*</td>
<td>.458</td>
<td>.480</td>
<td>3.07</td>
<td>0.458</td>
</tr>
<tr>
<td>IN</td>
<td>.245</td>
<td>.479</td>
<td>.401</td>
<td>.278</td>
<td>.287</td>
<td>3.5</td>
<td>0.681</td>
</tr>
<tr>
<td>M</td>
<td>3.77</td>
<td>3.93</td>
<td>3.87</td>
<td>4.07</td>
<td>4.03</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>0.961</td>
<td>0.821</td>
<td>0.915</td>
<td>0.942</td>
<td>0.64</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*p<0.05,  **p<0.01. P1—Sharing/Comparing, P2—Dissonance, P3—Negotiation/Co-construction, P4—Testing Tentative Constructions, P5—Statement/Application of Newly-Constructed Knowledge.
3.2 The Comparison of Interaction Performance between Native Speakers and Non-native Speakers

To compare the interaction performance of native speakers and non-native speakers, a t-test was implemented. The results are shown in Table 2. The mean value of interaction performance of the non-native speakers was higher than the mean value of native speakers in each phase. But all of the p value was more than 0.05. Although the interaction performance of the non-native speakers was better than that of native speakers, there was no significant difference between the two groups. Thus, the results indicate that the different native languages have no significant effect on student’s interaction performance in CSCL.

Table 2: Independent Samples Test.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Non-native speaker</th>
<th>Native speaker</th>
<th>t</th>
<th>p</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>4.21</td>
<td>.636</td>
<td>3.38</td>
<td>1.061</td>
<td>1.822</td>
</tr>
<tr>
<td>P2</td>
<td>4.14</td>
<td>.690</td>
<td>3.75</td>
<td>.926</td>
<td>.920</td>
</tr>
<tr>
<td>P3</td>
<td>4.21</td>
<td>.488</td>
<td>3.56</td>
<td>1.116</td>
<td>1.496</td>
</tr>
<tr>
<td>P4</td>
<td>4.50</td>
<td>.408</td>
<td>3.69</td>
<td>1.132</td>
<td>1.894</td>
</tr>
<tr>
<td>P5</td>
<td>4.29</td>
<td>.488</td>
<td>3.81</td>
<td>.704</td>
<td>1.490</td>
</tr>
</tbody>
</table>

*p<0.05, **p<0.01.

4. Conclusion

This study designed two activities to investigate the correlation between the interaction performance and ICC as well as the difference of interaction performance between students with different native languages. The results show that Dissonance and Negotiation are the most important phases of CSCL because they correlated to many factors of students’ ICC. Thus, when designing intercultural activities in CSCL context, teachers should pay more attention to these two phases. And it is necessary to provide more appropriate scaffolding and guidance in Dissonance and Negotiation. Besides, there is no significant difference of the interaction performance in CSCL between native speakers and non-native speakers. It indicates that native language has no significant effect on the performance of collaboration. This study is limited by the small sample size. It is unclear whether there is any significant difference between students’ interaction performance in CSCL. In our further study, we hope expand the scale of sample size and explore the specific factors that influence the students’ ICC in the CSCL.

References

Collaborative Learning in Elementary Science Supported by Learning by Inquiry and Augmented Reality

Sie Wai CHEW1, I-Hsiu LIN1, Yin-Cheng HUANG1, Kinshuk2 & Nian-Shing CHEN1*

1Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan
2College of Information, The University of North Texas, USA
*nschen@mis.nsysu.edu.tw

Abstract: The significance of science in the development of technology is undeniable and had proven its importance in the advancement of today’s technology. Hence it had result in the strong emphasis of science in the educational field. Elementary science consists of having students learning and understanding scientific concepts and knowledge. With the teacher-centered teaching method commonly adopted in schools, it would be difficult for students to fully understand and apply the learnt concepts and knowledge in real life. This research designed an application that utilized collaborative learning in the learning process, allowing students to solve problems by discussing with peers through the strategy of learning by inquiry by working on hands-on activities in augmented campus environments. The results of this research found that collaborative learning supported by learning by inquiry and augmented reality technology had positive impact on the learning performance of students in the subject of science.

Keywords: Collaborative learning, learning by inquiry, augmented reality

1. Introduction

In recent years, the development of science and technology played an important role in providing limitless opportunity for the progression across all fields and industries. Researchers and educators had took noticed of its importance, and had made relevant changes on the subject of science in various levels of curriculum. While newer student-centered methods are being explored, most classrooms are still adopting the teacher-centered method whereby teachers share the knowledge of the subject by lecturing in the classroom (Murphy et al., 2012). Since science concepts are typically expressed in an abstract and descriptive manner, teacher-centered method would require students to visualize these concepts after they have comprehended what was delivered to them in the classroom (Meijer et al., 2006).

As one of the essential 21st century skills, collaborative learning has shifted the education spectrum from teacher-centered to student-centered (Kay et al., 2011; Mishra et al., 2011). Collaborative learning requires students to work in small groups, “mutually searching for understanding, solutions, or meanings, or creating a product” (Smith et al., 1992, p. 11). It had enabled students to explore new learning topics together, delegating tasks among the group, knowledge sharing and conducting discussions to enhance the understanding of the group members (Dillenbour, 1999; Smith, et al., 1992). Furthermore, past research has proven that for the subject of science, the learning by inquiry approach provides an excellent opportunity for students to solve problems, try out new ideas, and understand new scientific concepts more effectively (Ann Haefner et al., 2004; Roussou, 2004; Tamir, 1990). With the current technology development, augmented reality could assist students to visualize new scientific concepts to facilitate their understanding (Chiang et al., 2014; Lu & Liu, 2015; Hwang et al., 2016). In addition to that, augmented reality play a significant role in improving students’ learning engagement, and enhance their interest and understanding regarding the subject as augmented reality technology can portray concepts and ideas more clearly to students (Lu & Liu, 2015; Hwang et al., 2016).

With past research indicating that collaborative learning would enhance students’ academic performance, social and generic skills along with the benefit of learning by inquiry in learning science,
this research aimed to improve the learning process and enhance the learning performance of students in elementary science. This research designed an application to assist students in learning elementary science via collaborative learning, using the strategy of learning by inquiry and the technology of augmented reality. The research question for this research is “How does the use of inquiry support and augmented reality in collaborative learning affect students’ learning performance in science?”

2. Research Design

2.1 System Design

An application, ScienMon, was designed, operating on Android tablets and utilized augmented reality to assist in answer or task validation. ScienMon application was designed for students to collaborate with members of their group while learning about source of energy, transfer of energy and force, and electric circuits. Each group was equipped with a tablet along with a box of items required for the completion of various hands-on activities (including galvanometer, solar boards, wires, card boards, wheels, wheel axle, scissors, motor, buzzer, a set of gearwheels and gear belts). The application began with six pieces of puzzles hiding the image of the final assignment (i.e., building a solar energy car). In order to retrieve and complete their final assignment, students were required to complete two tasks assigned to disclose each piece of the puzzle. The first task for each puzzle consisted of a multiple-choice question about the aboriginal culture around the school with unlimited number of attempts allowed. After providing the correct answer to the question of the first task, students were required to scan the landmark mentioned in the question by using the camera function of the tablet. This function was built using augmented reality whereby the landmark itself were the marker. After completing the question of the first task and scanning the designated landmark, the second task of the puzzle would be disclosed which consist of questions or tasks on the source of energy, transfer of energy and force, or electric circuits (i.e., multiple choice questions, fill in the blanks, and hands on activity). Students could retrieve the answers of the questions by completing some hands-on activities. For example, for the task on “What is a parallel circuit?” the group would use items provided to build a parallel circuit and verify their work by using the tablet’s camera to scan their work. As the parts of the items provided are being marked, the application using augmented reality technology will identify each component, verifying the group’s work and provide them with their result instantaneously. For the second task of the puzzle, each group would be given one attempt to answer the question. If they are unable to answer the question correctly, the group would lose that piece of the puzzle. After completing all the tasks for one puzzle, the group would move on to completing the next puzzle.

2.2 Research Process

The participants of this research were students from two elementary schools (Grade 4 to Grade 6) in Kaohsiung, Taiwan, with 19 males and 20 females. The participants from both schools were gathered at one of the school located in the rural area. They were assigned into ten groups with a mixture of gender and school of origin. Before commencing with the research, students completed a pretest individually, with a total of 12 questions covering the topics of source of energy, transfer of energy and force, and electric circuits. Questions in the pretest and posttest were selected from past examination questions on these topics and were verified by teachers for their difficulty level and appropriateness for the students. Thereafter, an introduction session of the interface and functions of the application (ScienMon) and a briefing for the form of the activities expected ahead were conducted in order to familiarize students with the research. Each group was provided with a tablet and a box of items required for the completion of hands on activities. Students began the session by using ScienMon and moving around the school, seeking for the landmark or location. There were total six puzzles, each containing a cultural question with answers located at six different landmarks (Task 1), and six hands-on questions or tasks to be completed by each group (Task 2). After completing all the six puzzles, students were required to complete the final assignment, where they were required to use the knowledge they learnt to build a solar car individually with the materials provided. Finally, students were then required to complete a posttest, which was similar to the pretest. The total time of the experiment was five hours.
3. Results and Discussion

This research’s objective was to improve elementary students’ learning process and performance in science by using collaborative learning, supported by learning by inquiry strategy and augmented reality. A t-test was conducted on the scores of the pre-test and the posttest of the students for this study with results \( t(38) = 6.57 \) (\( p < .05 \)), indicating that there was significant difference between the results of the pre-test and the posttest score. With the average score of the posttest \( (= 8.69) \) being higher than the pre-test \( (= 7.13) \), it can be concluded that the students’ performance had improved and the system designed could assist students in their learning performance. It was observed that students were more excited and eager to participate in learning as compared to learning in the classroom. Students would take on responsibilities to help the group move forward, playing different roles to ensure that their group could progress. However, during collaborative learning, there would be free riders of the group who contribute the minimal effort to the group. Hence, for collaborative learning to be successful, individual accountability and responsibility should be considered. It was observed that students would work better in groups of two or three; with clear task for each rotated role they were assigned. For future research, it is suggested that the designed application could record and monitor each student’s activity and contribution during the learning process, and take necessary actions to encourage students to participate.

Acknowledgements

This study is supported by the Ministry of Science and Technology, Taiwan, under project numbers MOST 106-2511-S-110 -002 -MY3, MOST 104-2511-S-110 -009 -MY3 and MOST 104-2511-S-110 -007 -MY3.

References

Collaborative Inductive Problem Solving Using an ICT Tool in an Elementary Science Classroom

Hirokazu KAWANOa* & Shu MATSUURAb

a Setagaya Elementary School attached to Tokyo Gakugei University, Japan
b Faculty of Education, Tokyo Gakugei University, Japan
*hirokazu@u-gakugei.ac.jp

Abstract: This paper describes the effect of a real-time tool, MetaMoji ClassRoom, on collaborative discussions around the varied results of science experiments in an elementary school. The principle of our science education is to find materials and scientific questions from everyday experiences and investigate these questions with experiments and observations. Sharing the results of varied group experiments, the class attempts inductive inference to arrive at a generalized conclusion. The effects of the collaboration tool on students’ awareness of other groups’ results were suggested from their notes. Significantly, more texts referring to others’ results were found in the students’ notes when the experimental results were shared using the tool. We then measured the time process of commenting on the results of the group experiments, visualized by the collaboration tool. Comments fulfilling students’ discussions were indicated within 20 min. This efficiency of collaboration enables the classroom to take enough time to carry out inductive inference to arrive at a generalized conclusion based on the variety of experimental observations.

Keywords: collaborative learning, real-time collaboration tool, ICT-based science education, inductive inference

1. Introduction

The goal of our science education is to encourage students to participate in scientific practices and to develop scientific explanations for real daily life phenomena. Scientific explanations are developed through arguments. Likewise, science learning is shaped by experience and collaborative discussions (2016, Lehrer and Schauble). Also, for most individuals, learning takes place most effectively conducted via social interactions (1978, Vygotsky, 2003, Redish).

The educational principle of this study is to engage students in a scientific exploration of the daily life phenomena and facilitate their collaboration to make inductive inferences. To enhance the efficiency of such collaborative activities within a limited time of the classroom and to let every student participate in the discussions, we introduce a collaboration software. In this study, we first examine whether the students are better aware of other groups’ experiment results collected on the tablet. Then we analyze the time process of their exchange of comments using the tool.

2. Method

The tablet devices used were Apple iPad Air 16GB and Mini2 16GB. Every student was provided with a tablet. The application software introduced was ClassRoom (MetaMoji Corporation). The ClassRoom assembled and shared handwritten notes, characters, and images uploaded to the cloud server, allowing simultaneous accesses from more than ten classrooms of 40 students. To collect the experiment results on the tool, the teacher prepared a note that consisted of the teacher’s layer and the student’s layer which was divided into nine cell experiment groups. The student’s layer consisted of the data layer and the discussion layer to overlay comment texts on the objects in the data layer.
First, comparison of the whiteboard and the tablet as the collaboration space was carried out in two classes, in which fifth-grade students had activities on the “solubility of objects” from October to December 2015. The control classroom, which was equipped with a whiteboard, was attended by 17 males and 20 females. The experiment class was attended by 18 males and 19 females with Apple iPad. The ages of students are from 10 to 11. Second, the time process of commenting using the tool was recorded at a lecture on the “plant growth” in July 2016. Fifth-grade students (17 males and 17 females, at the ages from 10 to 11) attended the class.

3. Results and Discussions

3.1 Comparison between Whiteboard and Tablet

Figure 1. Sharing and discussing the experiment results. Left: experiment results were gathered in the collaborative tool, ClassRoom, on the tablet. Students look at the result by enlarging or reducing the images. Right: experiment results were put on a whiteboard, and the students look at these results.

Figure 1 compares the manner of inspecting experiment results gathered in the tool (left) to the whiteboard (right). The students of the classroom with the tool can make notes on what they find in their tablets on their desks, while the students of the whiteboard class were seen to gather and talk with each other in front of the whiteboard.

Table 1 shows the number of students who referred to the other groups’ results (collaborative reference) in the worksheets. By the 5% level $\chi^2$ test, the number of collaborative reference were significantly higher when using the ClassRoom app compared to the whiteboard ($\chi^2 (1, N = 71) = 4.058, p < .05$). This indicates that the interface of the app is beneficial for noticing the other groups’ results and comparing them with their results while enhancing the attitude of students toward discussing the similarities or differences among the results.

<table>
<thead>
<tr>
<th>Collaborative Reference [person]</th>
<th>Other Description [person]</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassRoom App</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>Whiteboard</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>Sum</td>
<td>37</td>
<td>34</td>
</tr>
</tbody>
</table>

3.2 Commenting on Experimental Results

Figure 2 left shows a snapshot of the display. The students pasted comment labels on the results. Label colors indicated their views as approval of the results, disapproval, and a question to the group. Figure 2 right shows the time development of the number of labels pasted for experiments on plant seed germination and the condition of growth. In both experiments, approximately linear developments were found within the first 10 min. For plant germination, the rate of labels increased; $\eta$ for the first 9 min was $\eta = 2.9 \text{ min}^{-1}$ and for the conditions of growth, the value of $\eta$ for the first 16 min was $\eta = 5.0 \text{ min}^{-1}$, indicating the students concentrated on inspecting and commenting on other groups’ results intensely.
On average, the students pasted $2.8 \pm 1.7$ times for plant germination and $3.7 \pm 1.7$ times for plant growth. The activity seemed saturated within approximately 20 min. These results indicate a notable efficiency both in paying attention to the many sample data obtained under shared questions and in the arguments on the real data.

![Figure 2](image)

*Figure 2. Left:* display of the experimental results and comment in *ClassRoom* display. *Right:* time development of the number of labels pasted in *ClassRoom* display.

4. Conclusions

We introduced the collaboration tool *MetaMoji ClassRoom* to the science class of an elementary school, to facilitate students' exchange and discussion of the results of experiments with various conditions. The proportion of discussion that included the other group's results was found to be significantly higher for students using the collaboration tool, indicating that the use of the tool was beneficial both for bringing the variety of results together and for accessing each other's results.

The arguments were recorded by superimposing comment labels onto the images of the results. The students were found to enter comments repeatedly within approximately 20 min. This is efficient as an activity of the discussion of experiments, which is followed by the discussion to arrive at a generalized conclusion within the class time.

Thus, this study suggests that use of collaboration tool is effective for gathering information and facilitating students' discussion of a variety of experiment results before they take time to engage in inductive inference in the classroom.

Acknowledgements

We thank MetaMoji Corp. for helpful discussions. This work was funded by the Ministry of Education, Culture, Sports, Science and Technology through the HATO projects of Tokyo Gakugei University, and partially supported by a Grant-in-Aid for Scientific Research (C) 15K00912 from the same ministry.

References


The Design of a Portfolio-Based Reading Conversation Platform

Tzu-Chao CHIEN*a, Zhi-Hong CHENb & Tak-Wai CHANC

aGraduate Institute of Network Learning Technology, National Central University, Taiwan
bGraduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan
* brian@cl.ncu.edu.tw

Abstract: This paper proposes a design for a socially-purposed learning portfolio, in which students' learning data are collected in a purposeful way in order to motivate them to engage in social interaction. Based on this idea, a conversation platform is introduced, where the student's records, comments, material produced and topics discussed are presented on a profile page to encourage reading-related discussion. Furthermore, the system provides tips adapted for each student to encourage them to join the reading-related conversation based on their individual reading taste and discussion status. The concept of the system design is introduced and opportunities and challenges for the socially-purposed learning portfolio are discussed.

Keywords: conversation groups for reading, reading discussion, computer-supported collaborative learning

1. Introduction

From the viewpoint of social constructivism, knowledge can be repeatedly constructed through interaction with other people having different cognitive structures and cultural backgrounds (Berger & Luckmann, 1967). The emphasis is on the importance of collaborative and experience-based learning. The learning portfolio is deemed an effective tool for reinforcing the power of collaborative and experiential learning, which can encourage self-direction and self-regulation. As defined by Paulson, Paulson, and Meyer: “A portfolio is a purposeful collection of student work that exhibits the student’s efforts, progress, and achievements in one or more areas.” (Paulson, Paulson, & Meyer, 1991). The learning portfolio is useful as a means of presenting the student’s learning status and demonstrating their character through the learning process. The rapid development of information communication technology (ICT) meant that learning portfolios can be applied in various ways to create more opportunities for learning. Their production offers more flexible and effective ways to collect and evaluate the student’s learning, which can be applied for both formative and summative assessment and to demonstrate the student’s growth and performance.

Reading is also an important way to learn. Learning by reading books is not only an individual activity, but can also involve social cognitive construction. Reading conversation activities provides students with the opportunity to reorganize and reflect upon what they have learned through social interaction. Prior studies on learning portfolios have been primarily concerned with their assessment aspects, with less attention is paid to their social interaction triggering aspects. The purpose of this study is to design a socially-purposed learning portfolio where the student’s learning records, process records and productions are collected and used primarily to trigger active social interaction. The effectiveness of the idea is demonstrated by designing a social reading conversation system in which the student’s reading records, comments, discussion records and productions are applied as catalysts to motivate their intention to initiate and engage in reading discussion. The following paragraphs will introduce the designed system in detail as well as the potential application of a socially-purposed learning portfolio.
2. Three Tiers of Reading Conversation Triggers

Figure 1 shows the three tiers of reading conversation incentive conditions. The foundational tier is face-to-face conversation, which is the most natural way for people to communicate where students talk about what they have read with each other, sharing thoughts about their favorite books and introducing the stories in the classroom. However, face-to-face reading conversation relies upon relationships in a real-world environment. In other words, the effectiveness of face-to-face triggering will diminish during online communication. To deal with this problem, the second tier for triggering reading conversation includes the personal reading portfolio in the form of the “individual bookshelf”. As shown in Figure 1, the individual bookshelf can be seen as a metaphor of a student’s reading portfolio. It includes that student’s daily reading records, personal comments and productions such as book recommendations (Chien, Chen, Ku, Ko, & Chan, 2015). Based on these profiles, the students’ favorites and consensus of opinions can be explored and counted by system algorithms. The detected consensus information can be collected and presented to each student as an incentive to encourage reading conversation. Specifically, the system recommends to each individual student a list of books which they have never read before, but have been liked by other people with similar tastes in reading. The system can also introduce students to a new conversation partner who has the similar reading experiences or reading tastes. The third tier includes the reading conversation group. The groups are comprised of several students who have the same reading interests or learning intention. Each group also produces their own learning portfolio, including a common favorite book list and discussion histories. As in the case above, students can explore and identify the reading groups they might wish to engage their conversation by checking on their group learning portfolios manually, or through recommendations provided by the system algorithms.

3. System Design

The designed reading conversation system provides basic reading discussion functions including the posting of articles and replies to others. Information on each student’s reading status is collected by the system through a book recording function in which students can record their book using the ISBN (International Standard Book Number). The system will automatically establish a portal page for each book which includes basic information about the book, including author(s), publisher, as well as a brief introduction. All discussions about this book are embedded in this page, so that students can see all topics related to the books read after being recorded. Based on these fundamental functions, some more particular functions are developed. Some characteristics of these functions are introduced separately in the following paragraphs.

In addition to book discussion pages, taking on a similar approach, the system also automatically establishes a reader profile page when the student first registers and logs in with a new account. The reader profile page can be seen as a specific collection of individual information related to the reading portfolio specifically designed for social purposes. The reader profile page provides information about the reader, including recently read books, discussion topics and recently added book-friends. In particular, the reader profile page also presents a pie-chart to illustrate reader’s book
taste, so that visitors can consider whether to send a book-friend invitation to this student. The same mechanism is used on the group page. The group reading profile page is presented publicly when people establish a new reading conversation group. Visitors can see some information about this group, such as the list of favorite books and discussion topics, and then they can send a request for membership (see Figure 2).

![Individual Profile Page](image1)
![Group Profile Page](image2)

**Figure 2.** Profile Page Connections for Reading Conversation.

To facilitate active social interaction, the system provides each student with individually adapted recommendations. When a student looks at an individual profile page, the sidebar of the page will recommend certain books or conversation partners based on their reading and discussion status. For example, each student can see the following sidebars: “you may like these books based on the books you read/your book-friends”, “these people are potentially good partners based on your reading preferences/your reading process/your reading discussion”, “you may like these conversation groups based on the books you like/your reading discussion”.

4. Conclusions

This paper introduces a potential approach for the creation of a socially-purposed learning portfolio. The learning portfolios are intended to trigger social interaction. This idea is applied in practice by designing a reading conversation platform, in which individual and group learning portfolios, such as reading records, comments given, related productions and discussion topics are specifically collected for social purposes. The system algorithms can provide adaptive tips for selecting books and discussion partners based on each student’s reading and discussion status. Future work can focus on the collection process for assembling the learning portfolio and the development of system algorithms which are involved with the incorporation of data mining techniques.

Acknowledgements

This study was funded by the National Science Council of Taiwan (104-2511-S-008-009-MY3) and by the Research Center for Science and Technology for Learning, National Central University, Taiwan.

References

Improving Reading Comprehension using the Cooperative Mind Mapping Summary Strategy

Gloria Yi-Ming KAO*, Ju-I CHENG & Chin-Chung TSAI

Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei 106, Taiwan, ROC
*gloriakao@gmail.com

Abstract: Becoming an effective reader requires cultivating habits of extensive reading and acquiring efficient reading strategies. This study investigated the reading comprehension performance of learners with three different reading summary strategies: "cooperative mind maps," "co-text," and "individual mind maps." Participants were 272 high school students. The result showed that the "cooperative mind map group" significantly outperformed the "individual mind map group," indicating that the group discussion process is essential to build deeper comprehensive understanding of the reading material, given the same condition of using the mind mapping tool to write summaries. Our major results showed that combining group discussion with a mind map as a visual summary writing tool is the most beneficial approach for improving students’ reading performance. In addition, through the summary construction process using the mind mapping tool rather than a simple text tool, students could better identify important concepts, think about relationships and hierarchies between concepts, and thus developed a better ability to organize and understand the reading content.

Keywords: Reading strategy, reading comprehension, mind map, cooperative learning, individual learning.

1. Introduction

Reading is the basis of all kinds of learning. However, becoming an effective reader requires cultivating habits of extensive reading and acquiring efficient reading strategies. “Reading strategies” play a crucial role in understanding the essence of reading material, collocating attention, and monitoring the reading processes (Duke & Pearson, 2002). According to the Programme for International Student Assessment (PISA), summarization is an important reading strategy that could be used to improve students’ reading performance (OECD, 2013). It can be used to organize information for a better understanding of the reading material and to precisely describe the essence of the text. However, this is an ability that most senior high school students in Taiwan lack, based on observation of their performance on the PISA test. Literature on reading summarization falls into two categories, textual (Jeong, 2011) and graphic (Chang, Sung, & Chen, 2002) summarization. Comparing the effect of the collaborative textual summarization strategy with direct instruction, Jeong (2011) found that summarization helped fourth graders better identify main ideas and write summaries. On the other hand, studies have also provided empirical results supporting the advantages of presenting summaries with graphical visualization. Chang, Sung, and Chen (2002) found that combining concept maps with textual summarization can enhance text summarization abilities and improve learning performance. Similarly, some studies adopted mind maps (Buzan, 1991), a special type of concept map, and obtained positive results in terms of improving reading comprehension performance (Brinkmann, 2003; Merchie & Van Keer, 2012).

In light of these studies, we intended to integrate textual and graphical summarization strategies into the reading process to compare their possible different effects on students’ reading comprehension. We also aimed to identify whether incorporating the mind mapping tool in group or individual reading activities would be more beneficial for students.
2. Study

This study aimed to compare the effects of different types of summarization tools (textual vs. graphical summarization) and the forms of classroom-reading activities (individual reading vs. group reading) at the same time, to understand their impacts on students’ reading performance. It is worth exploring whether the effect of graphical summarization strategy in the context of group-based discussion is the most efficient reading strategy for students.

2.1 Participants

A total of 272 vocational high school students (aged 16-18) in Taiwan participated in this study. The students were assigned to one of the three groups with different reading summary strategies: the “cooperative mind maps,” “co-text,” and “individual mind map” groups. As a result, cooperative mind map groups (N = 107), co-text groups (N = 83), and individual mind map groups (N = 82) were formed. For the two groups with a cooperative summary writing process (the cooperative mind maps group and the co-text group), students were further required to form smaller groups of 3 to 4 students for discussion purposes.

2.2 Procedure

The whole experiment was conducted for around 2 months with 45 minutes per week. In the first week, all participants were administered a pre-test on PISA and were then assigned to the Cooperative mind maps group, the Co-text group, or the Individual mind map group. In the following weeks, participants underwent the major reading and summary sessions. The Cooperative mind maps group and the Individual mind map group were trained to use the Xmind mind mapping tool. They then took part in a reading activity using mind maps to summarize the main ideas of the article on “Media Literacy and Right to Communicate.” The Co-text group took part in the same activity without using mind maps. In addition, the Cooperative mind maps group and the Co-text group were instructed to engage in group discussion and to contrast their own perspectives during the process. In the last week, all participants were administered a post-test based on PISA.

3. Preliminary Results

To understand the possible different influences of the reading summary strategy on learners’ reading comprehension ability, an ANCOVA was performed to examine the differences in the PISA post-test scores of the three groups, with the covariate variable being the PISA pre-test score. Before conducting ANCOVA statistical analyses, we needed to assess whether the correlation between the covariate and the dependent variable differed significantly among the groups. There was no significant interaction between the three groups and the PISA pre-test score ($F = 1.22, p = .30$), indicating that the assumption of homogeneity of the regression slopes was met. The ANCOVA result (Table 1) indicated that significant difference existed between the three groups on the PISA test ($F = 3.04, p = .05, \eta^2 = .02$). In particular, participants in the cooperative mind maps group (M = 3.97, SD = .06) scored significantly higher than those in the individual mind map group (M = 3.73, SD = .07).

Table 1: Mean, standard deviation, and adjusted means for the PISA scores.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Group</th>
<th>N</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PISA score</td>
<td>1</td>
<td>107</td>
<td>3.66</td>
<td>.56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>83</td>
<td>3.51</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>82</td>
<td>3.55</td>
<td>.52</td>
</tr>
</tbody>
</table>

Group 1: Cooperative mind maps group; 2: Co-text group; 3: Individual mind map group
4. Conclusion and Discussion

Our results showed that the Cooperative mind maps group outperformed the Individual mind map group in the PISA test, suggesting that the mind map alone could not produce satisfactory reading comprehension. Using a mind map to present a reading summary first helps students to identify the key concepts and clarify the relationships among major concepts. When it is further implemented in a cooperative setting, the effect could be optimized to enhance students’ reading comprehension. Cooperative mind maps could help prompt students to focus on discussing the reading-related tasks rather than being distracted by aimless chatting during group discussion; they were prompted to collaboratively reflect on and examine the relationships between the concepts retrieved from the text.

Our preliminary results suggested that when teachers ask students to read, they should encourage them to adopt the mind mapping tool to turn their reading notes into graphic form, discuss with their classmates in a group-based setting, and then generate a final mind map to represent their final reading summary.

References


Resource Description Framework (RDF) Models for Representing the Revision Process in Research Support Systems

Harriet Nyanchama OCHARO\textsuperscript{a} & Shinobu HASEGAWA\textsuperscript{b}
\textsuperscript{a}School of Information Science, Japan Advanced Institute of Science and Technology, Japan
\textsuperscript{b}Research Center for Advanced Computing Infrastructure, Japan Advanced Institute of Science and Technology, Japan
*harriet.ocharo@jaist.ac.jp

Abstract: A general research support system essentially provides support for the daily research activities of students in higher education. One of the important research activities is the revision process that students go through to improve the quality of their research output after they receive feedback from their supervisors or reviewers. This concept paper presents the rationale for using the Resource Description Framework (RDF) to describe the revision process metadata in the form of a universal model that enables integration, interchange of information and standardized communication across various research support tools, processing and storage. With RDF, we can not only query the original metadata to find out the progress of the revision process, but we can also apply inference queries to discover relationships or patterns related to the overall research process. The inference queries will also enable students to reflect on and observe the changes the research drafts go through during revision leading to the final version. Such inferences cannot be realized when using a traditional relation database (RDB) or when using native file applications such as word processors. Using RDF, the logging and history of the revision process during daily research can be obtained for reflection, which is an important part of learning how to carry out research better. The design is currently in the process of being implemented and tested.

Keywords: RDF-based models, inference querying, revision process, research support system

1. Introduction

Students in universities and other institutions of higher education need a system to manage their daily research activities. Daily research activities include logging and tracking of the research activity process step by step, communication with supervisors, managing the research schedule etc. Also, a research support system helps the students with finding the relevant information, choosing the right tools and producing an effective presentation of research results (Yao, 2003). Tracking of research progress would also enable students to complete their research in time (Suhaily, Rozainun, & Azmi, 2015).

A typical research process is commonly divided into phases, such as problem definition, literature review, design, experimentation, testing and evaluation and presentation of results. There are currently many tools available to support various research activities. For example, there are reading tools that support the literature review process, and word processors that support the writing of research articles. However, there is limited research in a common framework that brings together the various sub-systems to support the common goal of research (Yao, 2003). A general goal of our research is to utilize a universal framework to provide research support throughout the entire research process.

One of the challenges when designing such a universal system to support the research activities of students is how to handle the various files produced as research output data to support the revision process in the system (Ocharo & Hasegawa, 2016). Depending on the field of research and the research phase, files or applications of different types may be produced. These files or applications may be unstructured, semi-structured or structured and this might present challenges when developing the system because of the need to write different routines to handle each file or application type.
In summary, there are two problems we encounter when developing a research support system. One is that the research output files are of different types depending on the research field and research phase. Thus it is difficult for a student who is trying to evaluate their overall research progress because of disjointed subsystems/tools that support different phases of research. It is a challenge to obtain information such as tracing a common link from the beginning to the end of research, calculating the overall time taken at each phase in comparison to other phases, total time taken, etc. However, using a common framework for logging research output files’ metadata such as the Resource Description Framework, a W3C recommendation, can overcome this challenge.

The second problem is how to improve support for the revision process as there are limitations to the software tools or systems that students use when creating or editing their research output. For example, students would like to find out information such as the number and type of unresolved comments, what common mistakes they might be making, what comments take the longest time to resolve and why, etc. With RDF, we cannot only query the metadata of the output files to obtain this information, but we can also apply inference queries to discover relationships or patterns related to the overall research process. Such inferences cannot be realized using a traditional RDB (relational database) or by using native file applications such as word processors or spreadsheets. This is the originality of our contribution to the field of research about research support systems for students in higher education.

In this paper, we propose using RDF to overcome the challenges of handling file types in a universal framework to support the revision process for all the phases of research. RDF uses a linking structure called a triple to describe two resources and the relationship between them. This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources. We seek to represent the metadata about the various research output files in RDF, while maintaining a link to the original file for reference. This is because file metadata is the main input needed to support the daily research activities of students, which include logging, tracking, scheduling, communication with the supervisor, and most importantly, revising the output file to improve its quality depending on the comments from supervisors. When the student needs to work on the original file, they follow the URI (Uniform Resource Identifier) to access it with a suitable application. In this way, the research system can support any kind of files produced during research, because the metadata is described using a common XML(Extended Markup Language)-based RDF model which enables integration, interchange of information with other applications and standardized communication, processing and storage. The idea is to create an RDF-XML modeling environment as a service layer that sits between the application logic layer and the data storage layer. This layer will create suitable metadata models in RDF-XML that enables the interchange of information between the application and data storage layers, as well as facilitating exchange of information with any other applications outside of the research support system. This will make it easier to develop and extend applications that support revision during the entire activity research process, and will also make it possible to discover useful patterns using inference queries.

The rest of the paper is structured as follows – section 2 is an overview of the scientific research process and revision process, where we also identify the requirements of a system to support the revision process. Section 3 is a review of the related literature. In section 4, we discuss the rationale for RDB. Section 5 is the conclusion and future work.

2. Research Activity Process and Revision Process

2.1 The Research Activity Process

We base our research activity process on the research process model proposed by (Fankfort-Nachmias Nachmias, 1992). Most scientific research follows a more or less similar process (Lynch, 2013). It starts with the problem definition phase, where the research theme is set. This is followed by a literature review phase, which is in turn followed by the design of experiment or model phase, the development or experimentation phase, the testing and evaluation phase and the presentation of results phase, in that order. We abstract the basic scientific research process in an iterative waterfall model as shown in Figure 1 which represents an instance of research activity that might be typical & sequential but not
necessarily the case for all research activities. The waterfall model implies that the researcher can always go back and revise any of the previous phases.

One of the important factors to consider when designing a system to support the daily research activities of students how to handle the various files produced as research output data. At different phases, students use different tools or subsystems to create research output. See Table 1 for examples of the research output from various phases and the corresponding tools used to generate the output by students in the information science research field. Handling each file type individually can be difficult as it will require building support for the many different file types that exist today. Our idea is to obtain the metadata about the files which includes a link to the original file location so that students can still use the native application to manipulate it. The metadata is what will be used to support the daily research activities of students.

![Diagram of research activity phases](image)

**Figure 1.** A typical representation of an instance of research activity

<table>
<thead>
<tr>
<th>Research Phase</th>
<th>Research Examples</th>
<th>Research Output</th>
<th>Description of output</th>
<th>Research subsystem/Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting a research theme</td>
<td>Research proposal</td>
<td>Text</td>
<td></td>
<td>Word processor e.g. Microsoft Word</td>
</tr>
<tr>
<td>Survey of related literature</td>
<td>Literature review chapter/doc</td>
<td>Text</td>
<td></td>
<td>Reading and citation manager e.g. Readcube, word processor</td>
</tr>
<tr>
<td>Design and modeling of system</td>
<td>System model</td>
<td>Uml diagrams, graphical illustrations</td>
<td></td>
<td>Graphics applications e.g. Microsoft Visio</td>
</tr>
<tr>
<td>Development</td>
<td>System prototype</td>
<td>Raw code, executable application, web application</td>
<td></td>
<td>Integrated Development Environments (IDEs) e.g. IDLE for python</td>
</tr>
<tr>
<td>Testing and Evaluation</td>
<td>Summarized results, visualizations</td>
<td>Statistics - text, graphs, tables</td>
<td></td>
<td>Statistical software packages e.g. SPSS</td>
</tr>
<tr>
<td>Presentation/publication of results</td>
<td>Presentation slides, thesis, journal paper, conference paper, poster</td>
<td>Text, PowerPoint slides</td>
<td></td>
<td>Word processor, Slide software</td>
</tr>
</tbody>
</table>

Table 1: Research phases and corresponding research output
2.2 The Revision Process

In section 2.1, we have discussed a general research process. However, depending on the field of research, laboratory style etc., there are variations in the actual process. The one aspect that is common to all is that students will produce some kind of research output, for example experiment results or files of code, and that they will receive feedback from their supervisor to improve the quality of this feedback. Hence all students undergo a revision process throughout their research. This especially applies to thesis/paper writing/presentation where students go through several drafts before arriving at a final copy. Presentation of research results is a very important stage that culminates the end of the research process. It is for this reason that our paper focuses on the revision process.

Even during daily research activities, students may take notes or undertake experiments to collect data, which they will eventually use to compile a report or some kind of output for a particular research phase. Once students produce research output, they undergo a revision process to improve the quality of their output after receiving feedback, in the form of comments for example, from their supervisor. Since a student’s research is evaluated based on their quality of research output, then the revision process is an important part of the research process.

There may be differences in what constitutes output from each research activity but any revision process can be described a revision model that describes the daily research activity followed by the uploading of an initial draft, feedback from a supervisor and subsequent revision by the student based on this feedback. This revision process repeats itself until the final draft is satisfactory. See Figure 2, which is adapted from Hasegawa & Yamane, (2011).

We identified the following self-describing metadata that we consider to be useful supporting the revision process: file author, file identifier, date created, due date, draft version, title, research phase, file description, file location (uniform resource identifier), and comments from the supervisor. The comments in turn contain more information such as comment identifier, comment author, date, due date, comment range, comment text, comment type and comment status.

![Figure 2. The research output revision process](image)

2.3 Requirements for Supporting the Revision Process

In order to provide logging, tracking, and support for the revision process throughout the research process, we need to integrate, store and manage data from all the different research subsystems. Figure 3 shows the interoperability needs with other research subsystems such as word processors, presentation tools, or data analytic tools. We have identified the requirements below to support the revision process:

- Ability to handle metadata from different file types while maintaining a link to the original file so that students can edit it in the native application
- Ability to handle unstructured, semi-structured or structured data from the many different subsystems as it is not possible to anticipate all file types
- Ability to query information about the revision process including the duration of revision, the due date of the research output, the number of unresolved comments, the author, total number of comments etc.
- Ability to perform inferential queries to obtain information such as what could be making the duration of the revision process too long, what could be the most difficult comments to resolve etc.
- Identification of common patterns emerging during research
- Knowledge transfer to new students in a laboratory in the form of a checklist of the most common comments

3. Related Literature

In the semantic web, information has well-defined meaning (Berners-Lee, Hendler, & Lassila, 2001). In other words, it can be thought of as a paradigm shift from a document-based web to a web of interlinked data (Open Data Support, 2014), where the data has meaning and relationships and is not just plain text. The semantic web is about enabling access to this data, by making it available in machine-readable formats and connecting it using URIs. The Resource Description Framework (RDF) is a W3C recommendation for representing data and resources on the web. In RDF, a resource is represented as a triple of a subject, predicate and object (W3Org, 2014). The subject is a URI identifying the resource, the object is another resource that the subject is related to, and the predicate describes the kind of relationship between the subject and object (Heath & Bizer, 2011).

There are several advantages of RDF discussed in the various literature. One important aspect of RDF is that the RDF data model is designed to enable integration of information that originates from multiple sources (Bergman, 2009). Information from different sources can be easily combined into a single graph. RDF data models allow for the representation of tightly structured data as well semi-structured data. It also enables the setting of RDF links between data from different sources (W3Org, 2014), which allows client applications to navigate between data sources and to discover additional data (Bizer, Heath, & Berners-Lee, 2009). Another advantage of using RDF models is the
ability to support inference queries. Inference queries allow us to generate or discover new relationships based on existing ones, such as is the case with logic programs. We can query RDF models using SPARQL, a query language for RDF. SPARQL can be used to express queries across diverse data sources (The SPARQL Working Group, 2013). SPARQL query results can be data sets or RDF graphs, which enable us to visualize the relationships in the data models.

One thing we have to consider when using RDF models to represent data in our research support system is linked open data. Open data is publicly available under an open license to be freely used, reused and redistributed (Open Data Support, 2014). Our aim is to develop a web system based on RDF-models and eventually make a contribution to the open linked data sets when the data models mature and are stable. As Bizer, Heath, & Berners-Lee (2009) state in their paper, the challenge now is for researchers and developers to create domain-specific applications that exploit the potential of linked (open) data. We have identified a research gap in RDF specifications for research support tools or systems. As Yao (2003) note in their research, there is a lack of study of such research support systems in a common framework. There are various research tools out there to support the various phases of research, so integrating all the data from the different tools into one overall system is a challenge that can be overcome by RDF-based models.

4. Why RDF and not RDB (Relational Database)?

There are several metadata formats available for describing data in web systems, including plain XML, HTML (Hypertext Markup Language) etc. The most common way to store data and metadata for general web applications, including research support tools, is in relational databases (RDB) and that is why in this section we focus on the discussion of the advantages of RDF over RDB. However, the advantages we discuss can be extended to cover the advantages RDF over all other metadata formats, particularly the ability to do semantic searches and inference queries.

We illustrate with an example of a student who is working on phase 1 of the research (see Table 1) where the expected output is a research proposal, typically a text file. The student may be using a word processor such as Microsoft Word to create and revise the document. However, when the student is revising, he or she cannot query Microsoft Word to obtain an overview, at a glance, of information such as:

- The total number of comments, the resolved comments, the open comments, persistent comments in subsequent drafts etc.
- The total time spent on revision so far, because there is no direct link between two distinct files that are separate drafts of the same document
- Commonly occurring comments in research proposals of other students so the student can know which common mistakes to look out for
- Other metadata about the research phase such as due date, document and comment authors, etc.

Our system’s aim is not to take over the functions of Microsoft Word but to act as a complementary tool in the revision process by utilizing research output file metadata. We could simply put the metadata in a relational database and query that. However, there are several reasons why we should consider an RDF-based application over a relational database (RDB) to fulfill the requirements of the revision system as discussed in section 2.3:

- RDF can handle metadata from any file type as it describes resources (research output) with the corresponding properties and property values. It can therefore handle metadata from different types of files whether structured or not. With RDBs, input data is always constrained. As it is not possible to anticipate all the different file types, marking up the data as RDF models is better.
- Interfacing or interchange of information with research support subsystems – for example if we require specific data in the research support system, we can automatically obtain information such as author, date etc. from some of the subsystems that have structured metadata. For instance, Microsoft Office Open XML (Ngo, E. C. M. A., TC45, 2008) is a zipped, XML-based file format developed by Microsoft for representing spreadsheets, charts, presentations and
word processing documents. With RDF-XML, it is easy to map these XML representations into RDF-based models. RDF-XML is a W3C standardized serialization format for RDF.

- When it comes to XML-RDF data models that facilitate interoperability, there are many open source and free tools to convert the models into various formats such as JSON, pdf, excel, csv, pdf etc. So it is possible to still have a conventional relational database to store the data or the data can be stored in a file-based format, and then mark-up that data into RDF models.

- Another reason is that nesting can be problematic for RDB especially when it comes to nested comments (feedback on research output). RDF is suitable for describing tree-like structures which are more intuitive for nested comments.

- With RDF, it is possible to use inference queries to identify common comments and similar patterns during research. Such inference queries are not possible with traditional RDB models. If we obtain RDF files from other students doing similar research, we can build an archive of knowledge and this can enable knowledge transfer to new students in a laboratory, for example in the form of a checklist of the most common comments for students to pre-check their research output.

Using our system, the new revision process can described as below (see Figure 4).

- The student will upload the document into the system, where the web application passes this document to the RDF modeling environment that maps the document metadata into RDF-XML model.

- The resulting file is then saved in the backend storage (can be any suitable storage system, the developer is free to choose). The data stored will be used for search and retrieval.

- The scheduler is then updated with the due date, and a notification is sent to the supervisor to review the uploaded document. The supervisor is also provided with a link to the location of the uploaded file.

- The supervisor can either directly add the comments to the word file, assuming it is a word file, or write their comments into the system.

- Once the supervisor has written the comments, the RDF model of the document metadata is updated as a 2nd version of the document, and a notification is sent to the student to improve the document based on the comments.

- Once the student revises their output and uploads the second version of the comment, the supervisor checks it and sets the status of the comments that have been adequately addressed as “closed”.

- If there are any new or unaddressed comments, the student is notified of them. This goes on until the statuses of all the comments are set to close, at which point the document is archived, along with the previous versions which will enable the student to reflect and learn from the revision process.

![Figure 4. Example of a Revision Process](image-url)
5. Conclusion and Future Work

In this concept paper, we presented the case for using RDF-based models to describe the metadata of research output files that can make it easier to develop and extend applications to support the daily research activities of students in higher education, especially the revision process that students go through to improve the quality of their output. RDF models can enable integration and exchange of information with the other existing subsystems that support various phases of research in a linked open data environment. Using RDF models enables us to collect all metadata from all different research phases, and using this data we can run inference queries to reflect and learn about the overall research process. We recommend using RDF over other metadata formats particularly because it is possible to do semantic searches and inference queries. Since research is a core activity of institutions of higher education, the importance of a research support system that aggregates data from various research in a common framework cannot be overestimated.

We are currently in the process of implementing and testing the usability and efficiency of the system. Future work will involve analyzing the research output metadata we collect using inference queries to discover common patterns in the research process, and sharing the lessons learned from inference querying. We eventually hope to make a contribution to the open linked data sets when the data models mature and are stable, thereby creating an open gateway for other developers or researchers to extend the system.

Acknowledgements

The work is supported in part by Grant-in-Aid for Scientific Research (C) (No. 17K00479) from the Ministry of Education, Science, and Culture of Japan.

References


Abstract: At the end of each course, students are required to give feedback on the course and instructor. This feedback includes quantitative rating using Likert scale and qualitative feedback as comments. Such qualitative feedback can provide valuable insights in helping the instructor enhance the course content and teaching delivery. However, the main challenge in analysing the qualitative feedback is the perceived increase in time and effort needed to manually process the textual comments. In this paper, we provide an automated solution for analysing comments, specifically extracting implicit suggestions from the students’ qualitative feedback comments. The implemented solution leverages existing text mining and data visualization techniques and comprises three stages namely data pre-processing, implicit suggestions extraction and visualization. We evaluated our solution using student feedback comments from seven undergraduate core courses taught at the School of Information Systems, Singapore Management University. The experiments show that the proposed solution generated suggestions from the comments with the F-Score of 78.1%.

Keywords: student feedback, teaching evaluation, implicit suggestions, text analytics, text mining, classification techniques

1. Introduction

Student feedback on course and instructor provide a wealth of information about students’ experiences in the course (Lizzio et al., 2002). Student evaluation systems help to counter anecdotal information about teaching behaviours and effectiveness. The evaluation systems provide a channel to systematically assess teaching and provide useful information about the effectiveness of teaching methods, instructor presentation, assignments, etc. (Moore & Kuol, 2005). This valuable feedback helps the instructor in improving the teaching and learning process (Murray, 1997; Elaine & Iain, 2005; Hounsell, 2003).

Education institutions conduct, analyse and disseminate the results of course evaluations either online, on paper, or through a combination of both methods. The student evaluation systems collect two types of evaluations; quantitative and qualitative. At the end of the course, the instructors will receive the evaluation reports for analysing their course delivery.

The quantitative evaluations are analysed, summarized and a table with statistics along with comparisons across the courses and faculty is generated for the instructors’ perusal. The qualitative evaluation report comprises the students’ feedback presented as a list of comments. The main challenge in analysing the qualitative feedback is the perceived increase in time and effort needed to assess written comments. For example, a single course, depending on the class size, can contain student comments that can range anywhere from 50 to 1000 sentences. Therefore, due to pragmatics, qualitative feedback from students is primarily conducted, evaluated and used for formative, rather than summative, purposes (Franklin, 2001; Lewis, 2001). Few research works explored how evaluations are used by several groups in the education such as instructors and administrators (Beran & Violato, 2007; Beran et al., 2005). The studies indicate that these groups rarely review written comments, and prefer use only what they perceive to be the more time-efficient, the quantitative ratings.

However, research undertaken by Harper and Kuh (2007) reveals that qualitative teaching evaluations can often bring to light issues that cannot emerge through conventional quantitative means. Beran et al. (2005) suggested that students, instructors and administrators ought to be offered training about the value of written comments and on techniques for, respectively, writing and analysing these
comments effectively. However, manually analysing qualitative feedback of large data is painstaking and tiring process due the high volumes of data. Therefore, there is a need for automated tools to analyse the qualitative comments and extract useful information from the comments and present insights in a user friendly format.

In general, the students' comments can be categorized into three types; Objective comments, Opinions and Suggestions.

**Objective comments:** An objective comment is a sentence which is completely unbiased. It is generally a fact about entities or events and their properties. For example “The programming fundamentals are taught in the first three weeks of the course” is an objective comment (Liu, 2010).

**Opinions:** Unlike factual information, opinions are subjective expressions that describe people’s sentiments and feelings towards entities or events (Beran et al., 2005). For example, “sometimes the instructor talks too fast for us to grasp the concept” is a sentiment towards the instructors’ presentation skills. A polarity can be assigned to the opinions. A single opinion from a single opinion holder is usually not sufficient for action. In most instances, one needs to analyse opinions from a large number of people,

- A **positive opinion** is usually a positive sentiment or feeling or likes of an opinion holder. For example, “the instructor is very knowledgeable, patient and easy-going”.

- A **negative opinion** is usually a negative sentiment or feeling or dislikes of an opinion holder. Negative opinions without context do not provide the details on the humans’ dislikes. For example, “I don't like this course” is a negative comment and the aspects of what he or she dislikes is unclear. Negative opinions with context are more useful. For example, “The project in this course is very heavy”, is a negative comment about the aspect, “project”.

**Suggestions:** Suggestions are comments that can be used for the product or service improvements (Ramanand et al, 2010; Brun & Hagege, 2013; Jhamtani, 2015). A suggestive comment in product reviews by a customers, aims to provide a suggestive intent for possible improvement of the product or service aspect. Figure 1 shows types of suggestions along with examples.

- **Implicit suggestions:** Implicit suggestions are expressed as wishes or improvements.

- **Explicit suggestions:** These are similar to the negative opinions. User likes and dislikes are taken into account to make recommendations. In the given example, “Sometimes he went through the concepts a bit too fast for us to grasp”, is a negative comments and one of the possible recommendation is that, “the instructor must slow his pace”.

![Figure 1. Types of suggestions from students’ comments](image)

In this paper we focus on extracting implicit suggestions from the students’ qualitative feedback using text mining approaches. There are several benefits of suggestions extraction task.
Firstly, suggestions are useful to improve the teaching and learning process in the course. For example, the suggestion such as “more programming examples should be given” is useful for enhancing the learning process of the students in the programming course. Secondly, when combined with the quantitative feedback, the suggestions help the instructor to prioritize and target the required changes that need to be applied to the course. Usually, the instructor uses the quantitative feedback on questions related to course and accordingly amends the course for improvements. In addition to quantitative feedback, the instructor can use suggestions which most students talk about and amend the course with more informed evidences on the specific components of the course. For example, if students provided a low score to “course labs, project and assignment” and then added suggestions in the comment sections, the instructor can use the suggestions on the labs, project and assignments to focus on the main issue and make relevant amendments. Therefore, the instructor can analyse where the main concern lies, whether it is in labs or projects or assignments, and amend the course accordingly. Thirdly, suggestions are useful to help improve the instructor's performance. We observe that the junior students take senior students’ advices in course bidding and selection. Hence, it is important for the instructor to improve his or her teaching and the course content. Through the course evaluation systems, the instructor has the opportunity to discover the gaps in the teaching delivery and course content. Applied in the effective manner, the instructor’s overall performance can be improved. Lastly, the management, dean or associated dean, can use the suggestions, to make decisions on the providing the necessary training or support to the instructor, for improving teaching delivery and course content.

One of the main challenges with implicit suggestions extraction task is the textual nature of comments which are expressed in natural language. We explain the challenges in detail in Section 2. Furthermore, the suggestions are embedded within the text. Opinion mining, topic extraction and NLP techniques (Liu, 2010; Sarawagi, 2008) from the text mining and linguistics research are widely popular for mining users’ comments in social media. Sentiment mining techniques are widely used for product review mining in consumer business world (Liu, 2010; Hu, 2004). We leverage these techniques for building the solution for implicit suggestion extraction task. Our solution applies data mining and text mining techniques on qualitative comments to extract suggestions from students’ comments.

Our work is novel in a way it focuses on implicit suggestions extraction from students’ comments. To the best of our knowledge there is no previous work that focuses on suggestions from students, using student qualitative feedback comments. We evaluated our solution using student feedback provided by the students for undergraduate core courses taught at the School of Information Systems, Singapore Management University collected over two semesters on seven different courses. Our experiments show that statistical classifier, decision tree C5.0 performs better than SVM with an overall F-Score of around 78.1% in extracting implicit suggestions task.

The paper will be structured as follows. Section 2 describes the implicit suggestion extraction task. Section 3 will be devoted to literature review on opinion mining research, suggestion extraction studies and student teaching evaluations research. Section 4 describes our implicit suggestion extraction solution overview and its details. In section 5, we focus on dataset, experiments, results and discussions. We conclude in Section 6 pointing some interesting future directions of our work.

2. Implicit Suggestion Extraction Task Definition

In this section, we explain the Implicit Suggestion Extraction task in detail. The input to the task is the list of student comments. Figure 2 shows sample inputs and outputs to our task.

Usually, the comments are short in nature but they may contain opinions as well as suggestions. For example, the first comment in Figure 2, contains an opinion as well as an implicit suggestion. “The course is good” is an opinion and “I do however feel that labs should be done in class to replace ICE” is an implicit suggestion. Also note that the fourth comment is a negative opinion with context and can be referred to as an explicit suggestion. In our work, we focus only on extracting the implicit suggestions from the students’ comments. The aim of the extraction task is to extract the sentences that are suggestions and provide the visuals to the faculty for deeper analysis.

The main challenges in the task include data challenges, suggestion identification and visuals dashboard generation.
In the next section, we provide some relevant literature review that provides the necessary background for our solution approach.

3. Literature Review

In this section, we provide some background on suggestion extraction and text mining algorithms. We rely on this background and leverage on some existing text mining techniques for designing our solution approach.

3.1 Opinion Mining

Opinion mining involves extracting sentiments and feeling from various sources like social media and online forum. Summarizing opinions helps government and businesses to adjust their governance policy and business strategy (Hu, 2004). Opinion mining has many real-life applications and several application-oriented research studies have been published (Liu, 2010; Sarawagi, 2008; Hu 2004). Opinion mining architecture takes users’ comments as inputs to generate sentiment analysis visualizations as outputs that can aid the decision makers in decision making process. Analysing student feedback focusses on automatically extracting opinions of students on course and instructor from large number of qualitative comments. In this work we leverage on the opinion architecture devised by Liu (Liu, 2010). Liu’s opinion architecture is built on text processing and text classification techniques.

3.2 Suggestion Extraction

Unlike opinion mining where we identify the like and dislikes or positive from negative comments. Extracting suggestions seeks to discover objective comments indicating what improvement an individual would like to see or have. Automatic discovery of suggestions from customer reviews or surveys is vital to understanding and addressing customer concerns. Equipped with this insight, businesses can channel their resources into improving their product or services. A previous study has employed rules based approach for identifying user wishes (Ramanand et al, 2010) through rule based method. There has been other research works in mining suggestion from sources like, tweets on mobile phone, electronics and hotel reviews (Sapna & Paul, 2015). Brun & Hagege developed a recommender system using customer profile and suggestions (Brun & Hagege, 2013). Yang et al. demonstrated that suggestion extraction can be applied in user recommendation based on user profile and geographical
context (Yang & Fang, 2013). In our work, we study the implicit suggestion extraction from the students’ course feedback. To our knowledge this is the first work in education data analytics research. We used classification based approaches for extracting implicit suggestions from qualitative comments.

### 3.3 Student Evaluations and Opinion Mining

Education institutions implement teaching evaluation surveys that enable comparisons to be made across the institution whilst allowing flexibility for individual course modules. These survey questions are the combination of “program-wide” questions and “module-specific” questions. Student surveys provide valuable feedback that helps course designers towards improving teaching style, course content and assessment design, and overall student learning (Murray, 1997; Elaine & Iain, 2005). Hounsell suggested that the feedback must be analysed and interpreted with great care so that action, and ultimately improvement, can result from feedback process (Hounsell, 2013).

Altrabsheh concluded that Support Vector Machines and complement Naïve Bayes produced the most accurate results while learning sentiment (Altrabsheh, 2014). They developed a system to analyse sentiments in real time to provide real-time intervention in the classroom. To predict whether or not a student would retake the course, Hajizadeh experimented on student feedback to analyse the sentiments (Hajizadeh, 2014). To study the opinion words from student feedback, Rashid used generalized sequential pattern mining and association rule mining (Rashid et al., 2013). Gottipati and Venky provided a framework for the qualitative feedback analysis (Gottipati & Venky, 2017). In this framework, they mentioned the challenges of textual data. Nitin et al, suggested a text mining approach to study the sentiments and topics in students’ comments (Nitin et al., 2015). However, the goal of all these previous works was to extract the sentiments and not the suggestions from the student feedback comments.

In our work, we are focusing on implicit suggestions extraction from students’ comments. To the best of our knowledge there is no previous work that focuses on suggestions extraction from student qualitative feedback comments. In our work, we leverage on some of the approaches suggested by the earlier works. We apply classification techniques in extracting suggestions from student qualitative feedback.

### 4. Solution

In this section, we first present the overview of our solution and then the details of each component of the solution.

#### 4.1 Solution Overview

Figure 3 shows the overview of our solution approach for implicit suggestion extraction. The solution approach consists of three main stages. In the first stage, raw input comments are pre-processed and prepared for suggestion extraction stage. The second stage is critical to our solution approach. This stage employs text mining algorithms for the extraction of suggestions from the processed comments. In the final stage, the extracted suggestions are aggregated for comprehensive reporting that can used by the instructors and administrators of improving the teaching and learning process.

![Figure 3. Solution approach for implicit suggestion extraction task](image-url)
4.2 Solution Details

In the first stage, to pre-process the data, all sentences are extracted from input comments using sentence tokenizers (Kurt, 2016). Tokenization deals with the splitting of text into units during data pre-processing which is critical for the second stage algorithms. Text can be tokenized in to paragraphs, sentences, phrases and single words. We also adopt a vector space representation of a document where each comment is evaluated as document term frequency.

Second stage involves extracting implicit suggestions using text classification methods. In our experiments, we used four different classification algorithms.

- Decision tree C5.0 - C5.0 is a classification tree that is produced by algorithms that identify various ways of splitting data. It is used to predict outcome or class to which the data belongs. A data is one comment and the outcome is either “a suggestion” or “not a suggestion” [27].
- Conditional Inference Tree (cTree) - cTrees work much like C5.0 decision tree for classification tasks. However, it uses significance test procedures to select variable and maximizing information measures [23].
- Generalized Linear Models (GLM) - GLM works on a fundamental principle of linear regression, an approach to model a relationship between variables used for prediction analysis. Each predictor has a coefficient with an assign level of significance or correlation to a certain class. The class in this case is “suggestion” or “not a suggestion” [24].
- Support vector machine (SVM) – SVMs finds a hyperplane that categorize the comments by their features over a space. The goal is to maximize the distance between the planes and points that falls on the edge of the plane known as support vectors for better performance of the algorithm [25, 26].

In last stage, the goal is to provide user friendly summaries of the suggestions obtained from student comments. The visualization interfaces design goal is to be more user friendly for search, comparison and analysis (see Figure 4). Graphical representation of the text using word clouds for a quick view is the most commonly used visual model. Additionally, we also designed query based tables style suggestions for better usability. We depict sample screen from our dashboard in Section 5.3. Other designs include the bar chart comparisons of the number of suggestions for various aspects of the course. The frequency of the suggestions will also enable the instructor to prioritize the changes that need to be undertaken for the course content or course delivery improvements.

5. Experiments

In this section, we first explain our datasets followed by results and discussions. Our experiments are designed to evaluate the implicit suggestion extraction stage and the visualization stage.

5.1 Datasets and Data Preparation

The dataset is the qualitative teaching evaluation feedback submitted by students attending undergraduate core courses offered by the School of Information Systems at Singapore Management University for two terms in a year. Not all comments are useful for analysing. For example, comments such as “NA” and “Nil” are discarded as they introduce noise into the datasets. After clean up, we have a total of 5,342 comments for our experiments.

Data Preparation for Experiments: To evaluate various classification methods, we first randomly sampled a small dataset, then we manually labelled the comments that are suggestions and finally, tested various classification methods described in previous section. To compare the models, we used text evaluation measures; precision, recall and F-Score (Hu, 2004). Precision is the fraction of comments that are actually suggestions among the total number comments classified as suggestions. Recall is the fraction of actual suggestions that have been retrieved over total number of suggestions in all the student comments. F-Score is the harmonic mean of precision and recall.
We took a subset of 399 comments randomly, to perform training and testing. We first perform sentence tokenizing (Kurt, 2016) on each of the 399 comment, which produces 604 sentences. For example, “Enthusiastic and entertaining. Classes were never boring. More in class exercises would be good to have” is tokenized into three sentences. We asked human judges to manually label these sentences as suggestions or not. We present the results in the next section and the visualizations of the tool in section 5.3.

5.2 Implicit Suggestion Classification Results

We evaluated four different statistical classifiers and the results are depicted in Table 1. We observe SVM and C5.0 give high precision and recall scores. C5.0 gives highest F-Score of 78.1% compared to other models.

Table 1: Evaluation results using different classification methods.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Linear Models (GLM)</td>
<td>0.676</td>
<td>0.650</td>
<td>0.658</td>
</tr>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>0.755</td>
<td>0.719</td>
<td>0.735</td>
</tr>
<tr>
<td>Conditional Inference Tree (CTREE)</td>
<td>0.781</td>
<td>0.681</td>
<td>0.698</td>
</tr>
<tr>
<td>Decision Tree C5.0</td>
<td>0.802</td>
<td>0.775</td>
<td>0.781</td>
</tr>
</tbody>
</table>

We further manually analysed the results to study the misclassifications that lead to low F-Scores. Table 2 shows some example comments and the predicted values by C5.0 classifier. We observe that the misclassified comments are poor in grammatical structure. One possible way of improving the tool performance is combining the rule-based or pattern-based techniques which can be considered as future work to improve this tool.

Table 2: Sample comments from the dataset and the predictions by the tool as implicit suggestion or not.

<table>
<thead>
<tr>
<th>Comment (Sentence Tokenized)</th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Prof could have given more leeway to teams seeking to enhance automation for clients.”</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>“We should have more practices in class to allow us to learn more stuff.”</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>“Lessons can be more engaging, by asking the students questions or trying out models.”</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>“Course could have spent more time on app logger and less time on the rest of the stuff”</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>“He tries to make the lessons as structured as he can.”</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>“Prior to this course, i never knew that Excel could be used to analyse or project future sales.”</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Probably organize lab sessions once a week for students to clear their doubts when they are using excel.</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Spends more time going through the examples as some students take more time to understand</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

5.3 Visualizations

For developing visual dashboard Shiny, a web application development tool is used [27]. A sample screen of the dashboard is depicted in Figure 4. Top of the screen provides keyword search functionality and a frequency bar to filter the suggestions. To the left of the screen, list of suggestions are presented in a tabular form and on the right is a word cloud. The word cloud gives an aerial perspective of the suggestion data, words that are of importance are highlighted by their size and colour. The instructor can use the word clouds as a reference to further refine their search and specify the frequency count for filtering the suggestions. For example, if the instructor would like to know what suggestions are given
for the word “assignment” which is highlighted in pink, he or she can enter this term in the search entry box.

In this example, we observe that students provide suggestions relating to the word “assignment.” Some sample suggestions include “assignment to be done in groups,” “provide clear objectives or direction” and “assignments to be in chunk size”.

![Figure 4. Visual dashboard for suggestions in a tabular form and word clouds. The tool enables to study a single course or compare across courses.](image)

6. Conclusion

Student suggestions aid the teaching faculty in improving the teaching content and class delivery for improving the learning process. In this paper, we proposed a new task of extracting implicit suggestions from students’ qualitative feedback comments. We adopted techniques based on text mining and opinion mining research. We observed that decision tree C5.0 classifier provides better performance with F-Score of 78.1%. Our future works includes detecting explicit suggestion where within a negative sentiment statement a suggestion is embedded. Future work will focus on extracting the topics within a suggestion, for example, instructor presentation, project work and exam. This would provide specific insight on what are the areas of improvement and highlights main concern within those suggestions. We are working on the further refining the visualization aspect of the dashboard based on feedback from instructors.

Acknowledgements

This research was supported by the Singapore Ministry of Education Tertiary Education Research Fund under the research grant reference number MOE2016-2-TR44. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore Ministry of Education. This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its International Research Centres in Singapore Funding Initiative.

References


Elaine Keane and Iain Mac Labhrainn (2005). “Obtaining Student Feedback on Teaching & Course Quality.” Centre for Excellence in Learning & Teaching. Apr 2005


Winston Chang (2016). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.5.3. https://CRAN.R-project.org/package=shinydashboard
Semi-Discovery Learning Support System for Analogical Reasoning in High-School Physics

Yasuhiro NOGUCHIa*, Takeshi ENOKIDAb, Tatsuhiro KONISHIa* & Yukihiro ITOHc

aFaculty of Informatics, Shizuoka University, Japan
bGraduate School of Integrated Science and Technology, Shizuoka University, Japan
cShizuoka University, Japan
*noguchi@inf.shizuoka.ac.jp

Abstract: In this paper, we describe a learning support system that supports the learning process by combining the processes of analogical reasoning and discovery learning. In learning high-school physics, as a method to deepen the understanding of physical phenomena, learners focus on mapping knowledge for known phenomenon and target phenomenon. This similarities-based mapping of knowledge enables learners to discover the type of influence between the knowledge parameters. However, before learners focus on the similarities, they must develop correct hypotheses for each phenomenon. Moreover, it is difficult for learners to create/modify/justify their hypotheses. In addition, they sometimes struggle to discover similarities in these phenomena. Our proposed system provides such learners support functions to complete the learning process. For our preliminary evaluation, nine participants completed pre- and post-test questionnaires using the proposed system. The results revealed that the system helped learners identify the relationships among parameters and similarities/differences among phenomena. However, the system did not help learners who could not identify the appropriate formula by viewing the graph. Therefore, future studies should focus on the needs of such learners and provide inputs to improve the system.

Keywords: Discovery learning, analogical reasoning, tutoring system, high-school physics

1. Introduction

Analogical reasoning is a thinking method to acquire new knowledge (hereinafter: target knowledge, TK) by applying known knowledge (hereinafter: base knowledge, BK) to new problems and situations. In this method, it is necessary to choose appropriate BK for TK, based on the semantic/structural similarities between the BK and TK. Numerous studies have indicated the importance of analogical reasoning in education and have reported application cases in some areas (Gadgil, Nokes-Malach, & Chi, 2012; Jee et al., 2010; Kurtz & Gentner, 2013). Day and Hills (2010) indicated that analogical comparisons used not only similarities but also differences between BK and TK.

As for high school physics, to deepen the understanding of physical phenomena via analogical reasoning, learners focus on the behavior of phenomena of BK/TK and the relationships among the parameters of these phenomena. Subsequently, the learners discover the type of influence of parameters by changing some parameters. In the cases of vertical spring and simple pendulum movements, learners observe these phenomena via T-k graph and T-m graph for BK and T-g graph and T-l graph for TK. While changing parameters in these phenomena, learners discover the relationships between parameter m and l, and K and g.

Base Knowledge: Vertical Spring Pendulum Movement

Formula: \[ T = 2\pi \sqrt{\frac{m}{k}} \]  
Formula: \( T \): cycle, \( m \): mass, \( k \): spring constant

Graphs: T-k graph and T-m graph

Target Knowledge: Simple Pendulum Movement

Formula: \[ T = 2\pi \sqrt{\frac{l}{g}} \]  
Formula: \( T \): cycle, \( l \): length of code, \( g \): gravity
Graphs: T-g graph and T-l graph

Relationships among parameters: m and l, K and g

Therefore, in this study, a learning process that combines the processes of analogical reasoning and discovery learning was employed. Discovery learning is a learning method wherein learners acquire knowledge by repeatedly creating and verifying hypotheses based on their observations. In discovery learning, learners can discover the relevance of phenomena through their trials and errors. However, in experiment-based discovery learning, learners are required to collect substantial data by changing input and situation parameters. Moreover, if the parameters to be observed are invisible, it is difficult for learners to observe the influence of the other changed parameters. In addition, learners struggle to create and verify hypotheses based on observations.

Some studies have focused on simulating and visualizing physical phenomena for education (Jose, Akshay, & Bhavani, 2014; Kaufmann & Meyer, 2008; Nancheva & Stoyanov, 2005); however, the simulators in these do not provide sufficient support to struggling learners to develop their own hypotheses. Horiguchi, Hirashima, and Forbus (2012) suggested an error-based simulation that visually displays learners’ incorrect formula, enabling them to identify errors by observing the visualization. Takeuchi (2000) suggested an intelligent tutoring system that allows learners to apply the knowledge acquired from a simple problem situation to a complicated problem situation; however, this tutoring system does not support the application of knowledge to different phenomena.

In this study, we focused on the learning of high-school physics. We proposed semi-discovery learning process wherein a system supports learners to create/verify/justify their hypotheses to reduce difficulties. The proposed system provides a virtual laboratory that can not only simulate and visualize phenomena but also generate graphs for the relationships between parameters. In addition, the system enables learners to identify the type of influence of the parameters by changing some parameters, thereby increasing focus on the similarities and differences between the BK/TK. In this paper, we describe the semi-discovery learning process with our supporting system and evaluate whether the system enables learners to (i) resolve the problem, (ii) understand the relationships among parameters, and (iii) identify the similarities and differences between BK and TK.

2. Semi-Discovery Learning for High-School Physics

2.1 Learning Process in Semi-Discovery Learning

The proposed semi-discovery learning process combines analogical reasoning, which comprises four steps: (1) selecting the BK, (2) mapping the BK and TK, (3) verifying the mapping, and (4) identifying the similarities/differences, and discovery learning, wherein learners acquire a rule system by iterating hypotheses verification. In physics, learners can acquire a model for phenomena by identifying the relationships among the parameters of the phenomena. For instance, a model of “Equations of motion” is $F = ma$; learners who have acquired the model and relationships among the parameters can attempt to apply the same structure into other phenomena. Consequently, they can acquire relationships between BK and TK through these trials. However, some learners face difficulties in identifying the relationships among the parameters. Therefore, the proposed semi-discovery learning process is simplified by a learning support system. Figure 1 displays our framework for the semi-discovery learning process comprising the following two steps.
1. Understanding each phenomenon
Learners learn phenomena through discovery learning processes: observation, development, verification, and updating of hypotheses. The proposed system suggests phenomena parameters that the learners should focus on. Subsequently, the learners observe the changes in the parameters of the phenomena in our system’s virtual laboratory. Thereafter, they develop a hypothesis that appropriately describes the relationships among parameters. The system generates graphs and formulas based on their hypotheses, and the learners compare the graphs and formulas with the real phenomena. Thus, through a continuous cyclical process, the learners verify and update their hypotheses.

2. Mapping two phenomena
After obtaining all graphs and formulas based on the BK and TK, the learners map them. In particular, they understand the similarities between the BK and TK by identifying common parameters. Subsequently, the learners clarify the differences between two phenomena based on the similarities.

2.2 Supporting Functions for Semi-Discovery Learning
To realize a framework for semi-discovery learning in physics, the supporting system should perform the following functions.

(a) Virtual Laboratory
In discovery learning, learners should be able to observe the changes in phenomena with changes in the parameters. Furthermore, in addition to highlighting the visible aspects, the system should display the invisible aspects (e.g., speed).

(b) Automatic Graph Drawing
Although it is important to summarize the input/output results of formulas, it is difficult to collect the input/output for the BK/TK, and learners often struggle. Therefore, the system should automatically generate graphs and parameters.

(c) Templates for Formulas
Novice learners generally face difficulties in developing hypotheses using formulas from the very beginning. Therefore, the system should provide templates based on formulas from high-school physics textbooks.

(d) Graph Feature Analysis
The system should inform the learners about the differences between their hypotheses and the correct phenomena through a graphical representation of learners’ hypotheses and the correct formula. In addition, the system should highlight the following differences in the graphs.
- Differences in the slopes.
- Whether the slope passes through the origin.
· Mathematical differences (modulus, absolute values, exponent, log, and square roots) in the formulas.

(e) Advisement
In discovery learning processes, learners typically struggle with identifying the errors in their hypotheses and are unable to update their hypotheses to rectify the errors. Therefore, to support such learners, the system should provide suggestions that highlight the errors in the learners’ hypotheses and what type of knowledge is required to update the hypotheses. The proposed system supports the following advisements.
· Differences in the features of the graphs.
· Differences in the parameters of the graphs.

2.3 Supporting Strategies

Step 1. Observing the BK and TK
The virtual laboratory displays animations of the phenomena for learners to observe the movement of the phenomena; for instance, the display of the vertical sprint pendulum movement.

Step 2. Acquiring the BK and TK
In Step 2, the learners create hypotheses using the graphs and model formulas to develop formulas and identify relationships among the parameters of BK/Tk.

Step 2.1. Developing hypotheses
· Hypotheses of the graphs
The system automatically generates sample graphs with appropriate types and ranges of parameters. The learners select a suitable graph from these sample graphs to develop their hypotheses. For instance, the system generates T-k graph and T-m graph for the BK.
· Hypotheses of the formulas
The learners select a suitable formula template from those provided by the system. Thereafter, they operate the parameters and constants in the template.

Step 2.2. Confirming hypotheses
· Hypotheses of the graphs
The learners confirm their hypotheses by comparing the hypothesized and correct graphs in the BK/Tk.
· Hypotheses of the formulas
The learners confirm their hypotheses by comparing the hypothesized and correct formula graphs in the BK. In addition, the system indicates the difference between these graphs.

Step 2.3. Justifying the hypotheses
The learners who identify the errors in their hypotheses repeat Step 2.1. Thereafter, they create updated hypotheses of graphs and formulas.

Step 3. Identifying the similarities in the parameters of the BK and TK
Prior to this step, the learners have acquired the models and graphs of the BK and TK. Thus, the learners are prepared to identify the similar parameters to derive a variable. For instance, T is derived based on k in the BK and g in the TK. The system displays alternative graphs of the BK and TK; the learners choose two similar graphs from these alternative graphs.

Step 4. Identifying the differences in the parameters of the BK and TK
The learners choose appropriate differences from the following alternatives.
The types of the differences:
(1) Parameters of the phenomena.
(2) Type of devices realizing the phenomena.
(3) Increasing the number of devices realizing the phenomena.
(4) Decreasing the number of devices realizing the phenomena.
Perspectives that require the learners’ attention, such as mechanics, electromagnetics, and atomic science.
Parameter ranges of the phenomena.
Variables in the BK and constants in the TK perform the same functions.
Constants in the BK and variables in the TK perform the same functions.

3. System Overview

3.1 Databases

The system has three databases (DB): Phenomena DB, Mapping DB, and Templates DB. The Phenomena DB defines formulas, relationships among parameters, and type of graphs for each phenomenon. The Mapping DB defines similarities and differences between the BK and TK. The Templates DB defines formula templates that learners use in creating their formula hypotheses.

3.2 System Interface

Figure 2 presents the main interface of the system. The interface has phenomena animation areas for the BK and TK to observe similarities and differences in their parameters. The concept-mapping table indicates similarities, differences, and types of differences between the BK and TK; these items indicate that the learners learned in the process. In addition, the system displays the number of rest relationships that learners should discover to highlight their progress. The bottom of this interface has three buttons for the BK and TK: go to virtual laboratory, reset the simulation, and create hypotheses.

Figure 2. Main interface of the system
Figure 3. Interface of Virtual Laboratory.

Figure 4. Interface for creating graph hypotheses.

Figure 5. Interface for creating formula hypotheses.
Figure 3 shows the interface of the Virtual Laboratory, developed using physics.js (Natural Science, 2017; Endo, 2015). Learners can change the parameters of the devices in the device parameters area and observe the values of the parameters simulated by the device parameters. In addition, the system prepares some tabs for appropriate types of graphs based on the BK and TK. When learners click on the tabs, the system generates graphs calculated on the basis of the device parameters.

Figure 4 shows the interface for learners to create graph hypotheses. Learners select an appropriate graph form to represent the relationship between the parameters. The system suggests 14 types of graph templates. When the learners select a graph, the system displays the selected graph and the graph generated by the correct formula in the situation. Learners validate their graph hypotheses by comparing the two graphs.

Figure 5 shows the interface for learners to create formula hypotheses. At first, learners select the appropriate formula template based on the elements (e.g., a radical sign, a fraction, a modulus, and other frequently used elements) including/excluding their formula hypotheses. The left tab in Figure 5 presents two formula templates that correspond with the conditions. Learners choose the appropriate formula and set appropriate variables/constants in the situation. Subsequently, the system generates two graphs: one using the learners’ formula and the other using the correct formula in the situation. Learners validate their formula hypotheses by comparing the two graphs.

Figure 6 shows the interface for identifying similarities and differences between the BK and TK. If learners create the correct graph and formula hypotheses for the BK and TK, the system redirects to this interface. In this interface, learners can choose relationships between the parameters of the BK and TK. The system displays the graphs for the parameters, and suggests learners to identify the similar graphs between the BK and TK. If the learners choose the similar graphs of the BK and TK, the system updates the concept-mapping table for the parameters used in the graphs as working similar in the BK and TK. Thereafter, the system suggests learners to choose the differences in the similarities from eight types of candidates mentioned in section 2.3. If the learners can choose the correct type, the system updates the concept-mapping table.

4. Evaluation Experiment

4.1 Hypotheses and Evaluation Process

We proposed the following three hypotheses:
H1: The system enables learners to complete the problems.
H2: The system enables learners to understand the relationships among the parameters of each phenomenon.
H3: The system enables learners to understand the similarities and/or differences among phenomena.
A total of nine university students participated in the preliminary evaluation of the system. We analyzed the differences between the participants using the proposed system and those using only the virtual experimental laboratory based on physics.js (hereinafter PHYSICS). A pre- and post-questionnaire was given to the participants. We defined the test administered to the participants after using PHYSICS as the pre-test, and that after using the proposed system as the post-test. Thereafter, we counted the blank responses in the pre-test based on the respective learning environment (i.e., the proposed system or PHYSICS). Because the form of the graph depended on the formula representing a phenomenon, we presented the participants with some problems about the formula representing the BK or TK and relationships among parameters. In addition, to analyze whether the participants could identify the similar parameters in the BK and TK, we presented them with some problems to identify the type of similar parameters in each phenomena and the type of difference among the combinations.

We prepared the following types of problems for the pre- and post-test.

- Relationships among parameters in the BK
- Formula representing the TK
- Relationships among the parameters in the TK
- Similarities among the combination of parameters
- Differences among the combination of parameters

Table 1 presents the steps, and the time taken to complete each step in the evaluation process.

<table>
<thead>
<tr>
<th>Step (min)</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-questionnaire (3)</td>
</tr>
<tr>
<td>2</td>
<td>Instructions (5)</td>
</tr>
<tr>
<td>3</td>
<td>Exercises using PHYSICS (30)</td>
</tr>
<tr>
<td>4</td>
<td>Tutorial (15)</td>
</tr>
<tr>
<td>5</td>
<td>Pre-test (10)</td>
</tr>
<tr>
<td>6</td>
<td>Exercises using the proposed system (30)</td>
</tr>
<tr>
<td>7</td>
<td>Post-test (10)</td>
</tr>
<tr>
<td>8</td>
<td>Post-questionnaire</td>
</tr>
</tbody>
</table>

Finally, we compared the pre- and post-test responses of the participants to validate the aforementioned three hypotheses.

4.2 Results

In the pre-questionnaire, all of the participants responded that they had learned physics in their high school.

Regarding H1, the average number of blanks responses in the post-test showed a 2.56 decrease from that of the pre-test (Table 2). Thus, the participants acquired some kind of knowledge about phenomena without facing any severe struggle in completing the exercise using the proposed system; perhaps, the participants had a real feeling of acquiring knowledge from the exercise. However, this did not imply that the learners acquired the correct knowledge. Therefore, we discussed the contents of their responses to validate H2.
Table 2. Number of blank responses in the pre- and post-test.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula representing BK</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Relationships among parameters in BK</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Formula representing TK</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Relationships among parameters in TK</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Similarity among the combination of parameters</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Difference among the combination of parameters</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>Average (per participant)</td>
<td>3.0</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 3. Number of correct responses in the pre- and post-test.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Difference</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Relationships among parameters in BK</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>0.89</td>
</tr>
<tr>
<td>Relationships among parameters in TK</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>0.89</td>
</tr>
<tr>
<td>(B) Formula representing BK</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>0.67</td>
</tr>
<tr>
<td>Formula representing TK</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>0.89</td>
</tr>
<tr>
<td>(C) Similarity among the combination of parameters</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>0.89</td>
</tr>
<tr>
<td>Difference among the combination of parameters</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>0.56</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>45</td>
<td>43</td>
<td>4.78</td>
</tr>
</tbody>
</table>

Row (A) in Table 2 indicates that the number of correct responses for the relationships among the parameters in both the BK and TK increased to 8 (0.89 per participant on average) in the post-test. Row (B) in Table 2 suggests that the average number of correct responses for the formula representing BK increased to 6 (0.67 per participant on average) in the post-test that of the formula representing TK increased to 8 (0.89 per participant on average). Thus, an increasing for correct responses was observed. However, there were two participants who did not provide the correct formula representing BK. One of them used the proposed method and although the pre-questionnaire response was not correct, this participant acquired a part of the correct formula for the exercise using the proposed system. The other participant could not complete the exercise; this participant acquired the correct graph between parameters in the BK but could not identify the structure of the formula from the features of the graph. The system notifies learners of their incorrect hypothesis for a formula and after the learners accept the notification, it provides a graph and a table, listing the formulas used to create the graph. Although the participant used this function, the participant struggled, and therefore we could not confirm the effectiveness of the function. Thus, H2 is confirmed by the result that correct responses per participant in the post-test increased to 43 (4.77 per subject on average) from the pre-test. In addition, this indicates that the system does not provide sufficient support to learners who cannot guess the appropriate formula from the form of graphs.

According to row (C) in Table 3, the correct responses about similarities increased to 8 (0.89 per participant on average) in the post-test. All participants who answered correctly completed the exercise using the proposed system; thus, the system enables learners to understand the common/similar parameters among phenomena. The correct responses about differences increased to 5 (0.56 per participant on average), which is relatively low compared with the increase in similarities responses. Thus, this preliminary evaluation validates H3.

5. Conclusion

In this paper, we proposed a semi-discovery learning support system for learning high-school physics. The system supports that learners understand similarities and differences in phenomena. It provides virtual laboratory, automatic drawing function, and other functions for creating/verifying/justifying learners’ hypotheses. In our evaluation with nine subjects, our proposed system was useful for learners...
to identify the relationships among parameters and similarities/differences among phenomena. However, the results indicated that the supports of the system were not sufficient for learners who could not guess the appropriate formula from the form of graphs. Future studies should perform an in-depth analysis of such learners and update the design of semi-discovery learning using a support system.

References


Using Data Analytics for Discovering Library Resource Insights – Case from Singapore Management University

LU Ning1, SONG Rui1, Dina HENG Li Gwek1, Swapna GOTTIPATI1,2* & Aaron Tay2
School of Information Systems1, SMU Libraries2, Singapore Management University
*swapnag.aarontay@smu.edu.sg

Abstract: Library resources are critical in supporting teaching, research and learning processes. Several universities have employed online platforms and infrastructure for enabling the online services to students, faculty and staff. To provide efficient services by understanding and predicting user needs libraries are looking into the area of data analytics. Library analytics in Singapore Management University is the project committed to provide an interface for data-intensive project collaboration, while supporting one of the library’s key pillars on its commitment to collaborate on initiatives with SMU Communities and external groups. In this paper, we study the transaction logs for user behavior analysis that can aid library admin to make operational decisions. The main challenges include the data quality and enormous datasets. Our solution not only provides the approach to data cleaning process but also suggest better visualization techniques for the user dashboard. Our experiment shows that the data cleaning process was effective in producing the insights from the library usage and the visualization techniques are efficient to summarize the big data. We used the datasets from Singapore Management University for this project.

Keywords: Library resources, students’ e-resource usage, data analytics, visualization models, horizon graph techniques.

1. Introduction

Online university library resources have become integral part of several courses in education institutions for supporting education and research for all stakeholders. The key function of university library is to meet the information, research, and curriculum needs of its students, faculty and staff (Reitz, 2004; Roseroka, 2004). The main focus is to support and improve the teaching, research and learning processes (Okunu et al, 2011). Many universities have online platforms and infrastructure in place to provide relevant information and respond to the users in an effective manner (Opoku, 2011). Continuous management and evaluation of the university libraries use is critical to ensure that the library is meeting the expected goals of the University. At the same time, adjustments should be made where necessary for effective information service delivery (Roseroka, 2004; KOUFOGIANNAKIS, 2012).

The resources collection at the Li Ka Shing Library, Singapore Management University is designed to provide resources necessary to support teaching and research by the Schools of Business, Economics, Accountancy, Information Systems and Social Sciences; and, courses taken by undergraduate and postgraduate students in all schools. The collection is interdisciplinary and covers the broad fields of business, economics and commerce with special strengths in finance, trade, accounting, management and international business.

Over years, the online collection has increased to over 250k e-book and e-journals of 80k titles. Some critical operational decisions that the library admins needs to make regarding the online resources include retaining of the resources, subscribing to new resources and unsubscribing to the resources. These massive amounts of information lead to various challenges in managing library resources. Manually studying the usage of resources is painstaking and tedious process and hence in this paper we propose an automated approach of studying the usage of library resources and thus aid in evidence based decision making process.
Data analytics has become more popular in various industries for finding insights in the business and making evidence based decisions by the business users (Sean Kandel et al., 2012; Baker et al., 2009). The main aim of this project is to identify the patterns of how the university student population utilizes e-resources from their library. We devised novel solutions to distil information from millions of proxy log records. A data processing workflow and two visualization concepts – horizon chart and word cloud – has been proposed to analyse on numerous resource providers and user groups.

The remaining paper is structured as follows. Section 2 will be devoted to literature review on data analytics research in libraries. Section 3 describes our solution overview and its details. In section 4, we focus on dataset, experiments, results and discussions. We conclude in Section 5 pointing some interesting future directions of our work.

2. Literature Review

Enterprises rely on data analytics to gain insights about their customers, products, competitors, and partners to make evidence based decisions in the day to day operations (Sean Kandel et al., 2012). Recent research in education field is showing growing interest in applying data analytics on education datasets to discover insights and hidden patterns. Applying data analytics techniques in education is an emerging research field and also known as educational data mining (EDM). It involves development of methods and tools for making discoveries within the unique types of data from educational settings. The goal is to better understand the stakeholders of education and the learning settings, and to gain insights of educational phenomena (Baker et al., 2009).

Libraries play a key role in the education and research process. The major stakeholders of the library resources are students. In one of the recent surveys, Cenage Learning’s Student Engagement Survey, more than 51% of the students responded that they are at the library to use the online databases, indicating that a good portion of their research work is completed at the library (Spring, 2015). Several researchers performed usability study to understand how users choose e-resources and thus identify ways to improve the access to e-resources (Fry et al., 2011; Maryellen Allen, 2001; Heather Jeffcoat King and Catherine M. Jannik, 2005). The approach is by survey based on Likert scale and open-ended questions (Fry et al., 2011) or surveys, focus groups and task-based testing (Maryellen Allen, 2001; Heather Jeffcoat King and Catherine M. Jannik, 2005). Some works used digital traces to understand the user behaviour in accessing the library resources. Analysis of digital traces aids in understanding of user behaviour in great details. Two types of web analysis are common; web search engine log analysis and digital library systems log analysis (Agosti, 2011). Stephen et al used webservice transaction logs to study how users find information from library college websites (Stephen Asunka and others, 2009). The focus of their work is to discover the issues relating to content access, interface design and general functionality of the website. Niu et al used transaction logs and surveys to understand users’ search behaviour and their preferences and perceptions of the two systems. They used transaction log analysis to assess the scope and distribution of search queries, the use of search options, as well as query construction and refinement (Niu, 2014). Digital library system log analysis is based on transaction logs which are well-organized and explicitly described library collections. In our work we use transaction logs to discover the usage of the library resources by various schools and the trends of usage.

Finding insights from the data logs is very beneficial to the library managers, as “Information derived from such analysis can thus serve as valuable input into the redesign and refinement of the library’s information systems, content architecture and the website’s interface” (Asunka et al., 2009). They serve as a huge motivation to find insights and support the decision making for the librarians. Furthermore, the data models are designed on various dimensions. For example, data arranged in chronological order, each semester’s record was broken up into sections by month (Asunka et al., 2009) can aid in discovering the trends. Such data structure can also include more dimension like student user profile, for example their course of study and admission year.

Just by accessing the raw data file, it provides no constructive clues to even start the analysis. Enlighten by Jansen, “A three-stage process composed of data collection, preparation, and analysis is presented for transaction log analysis” (Jansen, 2006). We leverage the ideas from (Jansen, 2006) to prepare our datasets. Hence, a more programmatic way of identifying the domain and logging errors is needed.
Despite the plethora of studies on library analytics, most remain on the theoretical level. In particular, few provide insights on the linkage between library digital contents and user demographics, which an important missing jigsaw in providing diversified services to the end user. In our work, we not only study the transaction logs to understand the e-resources usage analysis of library users in an efficient manner but also present the findings in much user friendly manner to the library admin for ease analysis. Comprehensive time horizon analysis (i.e. horizon chart) and content level analysis (word clouds) in the dashboards in our project is novel to the library research area.

3. Data Overview

Library proxy logs are the key datasets for this project. These are the digital traces of online library users of SMU. In this project we focus only on student data. The personal information and student demographics are anonymised. The dataset captures all user requests to external databases of e-books. SMU online library website sample and e-book data is shown in Figure 1. We observe that the SMU library website is enabled with various features for library users for easy information search across various types of datasets; books, articles, journals and exam papers.

![Figure 1. eBook sample from SMU library and screen shot of the SMU online library website.](image)

The users leave digital traces whenever they use the system and this project uses the transaction logs. As with any datasets, this dataset also comes with many data quality challenges. The dataset should be pre-processed with various techniques to prepare for the visualizations and analysis. The below are the major data quality challenges in the dataset.

- **Duplicate requests** are defined as a user makes multiple identical requests with the same URL within 30-second time span. Such requests are supposedly caused by auto page refresh.

- **Web assets** are used for page rendering and display. Web assets do not help us understand user behavior as the requests are not generated by user and are database dependent.

- **Domain names** are not explicitly available from the proxy logs as all the users requests are directly to the external common e-resources page. They are embedded in the URLs. Domain names of the articles are critical to understand the type of the resource requested by the users.

- **E-book database names** are also not explicit in the logs and they are embedded in the URLs. To
generate the user friendly reports for library admin, the database names are critical.

In our solution, we handle these challenges using rules and pattern matching techniques. In next section we describe our solution approach with examples.

4. Solution Approach

In this section, we explain our solution model in detail. Figure 2 shows the solution design of library insights extraction. The three main stages in our solution approach are “data cleaning design process”, “horizon chart design process” and “word cloud design process”.

**Data cleaning design process:** The main tasks include domain, date and database names extraction from the web assets and URLs. The example domain name extraction is depicted in Figure 3. We apply pattern matching techniques to extract the domain names.

![Figure 2](https://libproxy.smu.edu.sg:443/connect?session=s5hCfC0YaE07a20k&qurl=http%3a%2f%2fwww.economist.com%2f%3fsa_campaign%3dbulk%2feiu%2fsingaporemanagementuniversity%2fblank)

![Figure 3](https://www.economist.com)

Similarly, rule matching techniques are applied to discover the e-book database names. Table 1 shows the sample e-book databases and domain names from our dataset.

**Horizon chart design process:** The popular technique for visualizing time series data and compare the datasets is horizon graph (Few, 2008). The horizon chart reduces space use by dividing the chart into bands and layering the bands to create a nested form. The rationale to utilise the horizon chart is that it overpowers the usefulness of simple trend line in comparing data over time for
many items within a category. The traditional line chart that squeeze all the curves on a single graph is not palatable when it comes to tens of databases. However, it is still necessary to put them on a same perspective to make logical comparisons. Eventually, horizon chart becomes the dominant alternative to this situation. Converting from the trend lines to colour coded horizon chart not only helps segmenting messy trend lines into clear and straightforward heat map rows, but also insulate each element to include the trend lines with abnormal value for better user analysis (J. Heer et al., 2008; Tableau, 2016).

<table>
<thead>
<tr>
<th>e-book database</th>
<th>Domain URL patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>LawNet</td>
<td>*.lawnet.sg</td>
</tr>
<tr>
<td>WestLaw</td>
<td><em>.westlaw.</em></td>
</tr>
<tr>
<td>Ebsco</td>
<td>*.ebscohost.com</td>
</tr>
<tr>
<td>My iLibrary</td>
<td>*.myilibrary.com</td>
</tr>
<tr>
<td>Ebrary</td>
<td>*.ebrary.com</td>
</tr>
</tbody>
</table>

**Word cloud design process:** Another approach to derive insights with the data in common log files is via text mining on the extracted user accessed contents. Word cloud is the most popular visuals as it best illustrates the most searched topics in an easily understandable way. This will help the librarians to understand what each school searches on and cater recommendations to them later on. Two aspects been chosen for aggregation is by school and by batch. The rationale behind this course of action is to find out what each school and each batch is searching for the most.

5. **Experiments**

In this section, we first explain our datasets followed by results and discussions. Our experiments are designed to evaluate the implicit suggestion extraction stage and the visualization stage.

5.1 **Data Processing Results**

The dataset contains 22,427 records, not only limited to full-time but also postgraduates and exchange students. The request log records are filed by months. The monthly numbers of records in request log data vary from 3 million to 11 million and the file sizes range between 1GB to 3GB. Requests are produced more in March, September and October, and few in school holidays. Figure 4 shows the statistics of the log requests in 2016.

![Figure 4. Number of log requests for each month in 2016](image)

As discussed in the previous section, the first task is the removal of duplicate request. The removed duplicate records amount to 5% of all proxy requests. Generally, the number of duplicate lines
correlates with the number of requests for each database. Only half of the requests are directed to web resources. In reality, page rendering typically requires many web resources files, which predicts a large percentage of web asset requests. The contradiction is explained by the fact that only a small proportion of non-web resources requests are requests to web pages loaded by browser. A large proportion of all requests are AJAX requests used by web pages to load parts of the web app interface.

5.2 Horizon Chart Analysis

Dashboard with filter enables the user to interactively filter and change the results being displayed. For instance, Figure 3 depicts the usage rate of e-library by ‘Undergraduates’ from ‘School of Information System’ and ‘School of Accountancy’ that are currently in their first year. “Daily percentage diff” is calculated from ‘daily usage rate’ of each domain to its own “yearly average usage rate”. The values are then represented by two colour schemes. The bluish colour shows the less proportionate daily usages rate to its yearly average and on the other hand, the reddish colour shows the more proportionate end. The deeper colour represents the higher differences of the values. The horizon diagram will then only display the relevant usage rate per the criteria being selected. Furthermore, to compare the difference between the most and third popular domain used, we can click on the ‘Domain’ colour list.

Figure 5. Usage Pattern Dashboard on 1st year Undergraduate from SIS and SOA

From Figure 5, we observe some similarities lies within each trend lines. The usage rate of online library platform starts to surge from week 7 onwards across all three schools, and plunges during none school terms. This is in line with the school academic schedules, as well as the course schedules where most projects only kick starts in the second half of the semester.

However, exceptions are also very prominent in this case, as the students from SOB start to generate more accesses from week 5 onwards, which is much earlier than the other two schools. It may lead by a different course structure in school of business as they have an interim report due in the first half of the semester. Even though the usage rate by students from SOB and SOSS has three similar period of spikes in the second half of the semesters, SOSS students demonstrate a more consistent high demand than SOB students as the bar constantly filled with warm colours. We also observe that the students from school of business and law tend to behave differently across their term of study. It is also true that the contents they access for may not always align with their course of study.
5.3 Word Clouds Analysis

Word clouds be generated on various dimensions. In our project, we collected data only about schools. Figure 6 shows word clouds aggregated by school/faculty and aids to find out the most popular search queries for different faculties.

![Word cloud of Business School](image1)

![Word cloud of Information Systems School](image2)

![Word cloud of Accountancy School](image3)

![Word cloud of Social Sciences School](image4)

**Figure 6. Word Cloud for Faculties**

From Figure 6, we observe that “Tourism” is a hot topic for various school. School of Business, Economics, Law and Social Sciences students search within their own major whereas School of Accountancy and Information Systems have the tendency to search outside their major. Its purpose is a preparation for the librarians to develop recommender system based on their search queries. This can also value add to the online library platform for the students by showing more relevant articles on top while they perform searching. Further, this also aids the library admin to study on subscriptions of e-books databases. This requires deeper analysis with more detailed datasets which will be our future work.

5.4 Discussions

In our solution, we adopted heuristic approaches to handle the data challenges to the specific task of extracting insights of library e-resources usage. However, they is a dire need for a conceptual library analytics framework that can be of more generalised and aid in various operational decision process in the library. Understanding the meaning of each URL is a challenging, if not impossible, task. Extracting content information from URLs that are generated by various content providers requires a large amount of manual work. However, programming can largely reduce the pain by sacrificing a small degree of accuracy for higher efficiency.

There are a few limitations in this study. Firstly, user contents are not always available from the URLs as many contents are encoded into database-specific IDs. This reduces the base of analysis and possibly led to biased results. Besides, few attributes of students were provided, this restrained in explaining certain usage patterns. If the registered courses were given, we could have obtained a more confident correlation between courses taken and e-resources accessed.

Instead of simply using trend line to represent the usage pattern of the e-databases, horizon chart that we built can give a clear overview on the library proxy log data. It is also able to interactively alter the results with various filters attached to the chart which enables it to discover new insights in the future. From the word cloud, the library can quickly recognise the most popular resources. It also helps the library to sieve out valuable information otherwise hard to identify from a pile of proxy logs. Word cloud and horizon chart implemented in this project are succinct and dense in information. These visualisations not only deliver useful information, but also serve as a proxy for users to interact with the data. The charts are still limited due to the type of data collected in our project. With additional data, the visualizations can be further improved.
6. Conclusion

The analysis on the proxy log data has revealed that the e-resources usage patterns from different groups of students are significantly different. Despite the ability to identify the user behaviours, we realised the courses are the key factor that influences e-resources usage patterns rather than students themselves. From another perspective, students are lacks the motivation to read on their own. Future study can be done on packaging the analysis pipeline and automating the process. This will enable the library admin team to dynamically generate reports or constantly monitor for resource anomaly. Such effort will greatly improve the timeliness and accuracy of the analysis at the point of request. Expanding the time dimension would produce trend chronological patterns in a longer time frame and how students’ behaviours have evolved as well as how their topics of interest have changed over time. Further, we are working on the conceptual framework for library analytics for deeper analysis of the user behaviour and aid operational decision making process of the admin from library.

References


Tableau horizon charts. https://www.tableau.com/learn/tutorials/on-demand/horizon-charts
An Educational Support System based on Automatic Impasse Detection in Programming Exercises

Koichi YAMASHITA*, Takumi SUGIYAMA, Satoru KOGURE, Yasuhiro NOGUCHI, Tatsuhiro KONISHI & Yukihiro ITOH

Faculty of Business Administration, Tokoha University, Japan
Graduate school of Integrated Science and Technology, Shizuoka University, Japan
Faculty of Informatics, Shizuoka University, Japan
*yamasita@hm.tokoha-u.ac.jp

Abstract: In this paper, we develop an educational support system based on an automatic impasse detection method. In programming education, novice learners occasionally experience various impasses during coding exercises. A learner’s impasse log, which includes information such as the type of the impasses and timestamps at which they reached the impasses, would be useful for the learning and teaching of programming, such as for reflective learning and one-on-one instructions. Although several systems have been developed to understand students’ programming statuses, there is no system based on the students’ impasse in programming. We developed a ruleset to detect impasses automatically based on the indications of students’ codes and errors and improved the detection ability through an experimental introduction in actual classrooms. Using this framework, we developed a system that supports learners to perform reflective learning. We conducted a pilot experiment to evaluate the contribution of our system to reflective learning and obtained positive evaluation results.

Keywords: Educational support system, automatic impasse detection, reflective learning

1. Introduction

In programming education, novice learners occasionally experience various impasses during coding exercises; they might be unable to identify correct implementations or resolve certain compilation errors for a long time, and therefore not be able to proceed with their exercise. However, in the typical style of programming exercises, teachers and teaching assistants tend to have a relatively large number of students to monitor and understand every student’s impasses individually and precisely.

We consider that, unlike students coding effortlessly, those reaching an impasse would have some indication during their coding process. If we can identify such indications of impasses by observing how novice learners perform their coding exercises and how they reach impasses, it could be feasible to generate a ruleset for impasse detection. Based on this consideration, it would be possible to implement an automatic, rule-based impasse detection and record it in an impasse log.

A learner’s impasse log, which includes the type of the impasse, a snapshot of the program-code, and the compilation and run-time errors at that time, can be useful for learning and teaching programming; specifically, we can exemplify student’s reflective learning, one-on-one instruction during the exercise, and classroom designing. During reflective learning, students can reflect how they had reached impasse and how they had broken it using their personal impasse log as a clue. During classroom exercise, teachers can locate the students reaching an impasse and identify the type of impasse by monitoring their impasse logs in real-time. For classroom designing, the impasse logs can focus the teachers’ attention on learning items that had been difficult for many students to understand.

In this paper, we describe a system that records impasse logs and visualizes the summary. To develop the ruleset for impasse detection, we observed some novice learners’ coding activities and whether the learners reached an impasse. Furthermore, we recorded the snapshot of their program-code and the compilation and run-time errors every a certain time period. Based on the codes and errors recorded at the time when the learners had reached an impasse, we generated a ruleset for impasse
detection. We describe the ruleset development in Section 3 and our system in Section 4. We introduced our system into an actual class of 110 students for three months, and collected nearly 2000 impasse logs. In Section 5, we describe the pilot experiment that we conducted to verify our system’s contribution to reflective learning. The evaluation results suggest that our system supports student’s reflection to a certain degree.

2. Related Work

The concept to support learning and education by monitoring learners’ behaviors has a long history in computing education research. The most intuitive approach is tracking learners’ activities using the logging data of course management systems (Mazza & Dimitrova, 2004) in a certain computer-based learning environment (Biswas & Sulcer, 2010). In the context of programming education, Thomas et al. (2003) monitored the generic computer usage of programming students, recording low level actions such as mouse clicks, typing, and window changes. They identified the abstracted meanings and purposes of students’ action by analyzing the recorded actions. However, additional research is required because it is very difficult to analyze large-scale data using a bottom-up approach and to extract meaningful actions from them.

Compiler error messages are frequently used to understand the status of students’ programming. Brown and Altadmri (2014) collected the error messages of a large number of students (more than 100,000) from numerous institutions and analyzed the frequently made programming mistakes of novice learners. They highlighted teachers’ subjective impressions about novice learners’ common mistakes. Hartmann, MacDougall, Brandt, and Klemmer (2010) developed the HelpMeOut system that supports learners to debug compilation errors. HelpMeOut tracks code evolution over time and collects learners’ modifications that take program-code from an error state to error-free state. When a learner experiences a compilation error, HelpMeOut, as a sample solution, suggests the modification of the other learner who had experienced the same compilation error in the past.

Piech et al. (2010) adopted an approach that uses the changes in learners’ program-code. To model how students learn programming, they collected students’ code at certain time intervals and at every time a student compiled a project. They developed a programming model of students by modeling student progress, based on a Hidden Markov Model. They estimated transition parameters using recoded codes as observed outputs. Their model could potentially provide insights into whether students require interventions.

Our study is novel because we supported learners to not proceed with their programming exercises, but to perform reflective learning. Our basic idea also includes supporting teachers to provide one-to-one instruction in the exercise, and to design classroom. On that point, our approach is similar to the study of Piech et al. However, our approach is different from it in that ours uses both learners’ codes and errors and is based on an exercise-independent ruleset. Moreover, our impasse detection is based on abstracted indications of learners’ coding activities, whereas that of Piech et al. (2010) is based on the superficial indications appearing on the learners’ codes.

3. Ruleset Development for Impasse Detection

3.1 Ruleset Development Based on Observing Learners’ Coding Activities

For impasse detection, we recognized the indications appearing only on learner’s program-code and compilation/run-time errors. Learners’ coding activities provide other information such as the points of their gaze or the motions of their eyes, captured using eye tracking devices. However, this information is not realistic because such devices might not be introduced into actual coding exercises normally, and hence they would impose a burden to the learners. An externalization of learner’s thinking is also difficult for a similar reason. Therefore, we developed a subsystem to record four types of learner information: every one-minute program-code, every compiled program-code, every (if any) compilation error, and every (if any) run-time error.
To collect instances of impasse indications appearing on these types of information, we invited five voluntary undergraduate students majoring in computer science. We gave them programming tasks and observed their coding activities. The task involved solving few problems from a past regional Association for Computing Machinery / International Collegiate Programming Contest (ACM-ICPC) that the students had never seen. The total number of observed students was 15 and the total time of observation was approximately 30 hours. During their problem-solving, we collected abovementioned information using our subsystem and observed whether the learners reached an impasse; if yes, what were the causes of the impasse.

Through the observations, we classified learners’ impasses into the following four types:

Type 1. Learners unable to develop implementation strategy (including requirements interpretations, algorithm and data-structures design, and program design).

Type 2. Learners unable to resolve compilation errors.

Type 3. Learners unable to resolve run-time errors.

Type 4. Learners unable to modify the code to function as they expected.

We investigated all recorded codes and errors on each impasse type. Table 1 presents the indications appearing on the codes and errors for each impasse type derived from the investigation.

<table>
<thead>
<tr>
<th>Type of Impasse</th>
<th>Indications</th>
<th>Conditions to Detect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>(1-1) The learner has not modified the code for ≥15 minutes.</td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>(2-1) Compilation errors do not decrease even when the learner has compiled successively.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2-2) Compilation errors once resolved have reoccurred.</td>
<td></td>
</tr>
<tr>
<td>Type 3</td>
<td>(3-1) Same run-time errors have occurred successively.</td>
<td></td>
</tr>
<tr>
<td>Type 4</td>
<td>(4-1) The learner has repeatedly modified the same position in the code.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4-2) Statement calling a standard output function, such as printf(), with some variable as arguments has been added and then removed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4-3) Compilation has been repeatedly executed in a short time period.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4-4) Code has been re-modified to a previous code.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4-5) A large part of code has been modified at once.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4-6) Statements including the same variable have been modified successively.</td>
<td></td>
</tr>
</tbody>
</table>

The automatic impasse detection could be performed by recording learners’ codes and errors in real time, scanning the indications (Table 1) in the recorded materials, and reporting corresponding type of impasse if an indication was found. Based on results of this investigation, we generated the ruleset for impasse detection, presented in Table 2, setting underlined threshold parameters. The following threshold parameters were derived from our investigation.

<table>
<thead>
<tr>
<th>Type of Impasse</th>
<th>Conditions to Detect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>(1-1) The learner has not modified the code for ≥15 minutes.</td>
</tr>
<tr>
<td>Type 2</td>
<td>(2-1) Compilation errors have not decreased between any two sequential compilations by the learner.</td>
</tr>
<tr>
<td></td>
<td>(2-2) Compilation errors once resolved have reoccurred.</td>
</tr>
<tr>
<td>Type 3</td>
<td>(3-1) Same run-time errors occurred ≥7 times, successively.</td>
</tr>
<tr>
<td>Type 4</td>
<td>(4-1) Same position in the code has been successively modified by the learner ≥2 times.</td>
</tr>
<tr>
<td></td>
<td>(4-2) Statement calling a standard output function, such as printf(), with some variable as arguments has been added and then removed.</td>
</tr>
<tr>
<td></td>
<td>(4-3) Compilation has been executed ≥2 times in ≤3 minutes.</td>
</tr>
<tr>
<td></td>
<td>(4-4) Code has been re-modified to a previous code.</td>
</tr>
<tr>
<td></td>
<td>(4-5) ≥15% of the code has been modified at once.</td>
</tr>
<tr>
<td></td>
<td>(4-6) Statements including the same variable have been modified ≥3 times, successively.</td>
</tr>
</tbody>
</table>
3.2 Improving the Detection Ability of the Ruleset

We introduced the subsystem of impasse detection into actual classes of an introductory Java programming course, from October 7, 2016, to November 17, 2016, (6 weeks). We aimed to evaluate the detection ability of the ruleset described in the previous subsection, and to adjust rules or threshold parameters to improve the detection ability. The course named “Programming” is offered to first grade undergraduate students majoring in computer science. A total of 110 students were enrolled in the course. The subsystem detected 2104 impasses within our experimental introduction, and recorded them into impasse logs, which included the impasse types and the snapshot of the codes and errors. To verify the validity of the ruleset, we measured the recall and precision of detected impasse logs as follows:

\[
\text{precision} = \frac{\text{N of correct detections}}{\text{N of correct detections} + \text{N of incorrect detections}}
\]

\[
\text{recall} = \frac{\text{N of correct detections}}{\text{N of impasses detected by hand from all recorded codes and errors}}
\]

In the precision calculation, each impasse of all 2104 detections was manually classified into correct, incorrect, or non-identifiable, judging from the codes and errors recorded with the impasse. As the result, we obtained 938 correct, 1004 incorrect, and 162 non-identifiable detections; hence, the precision was .483. In the recall calculation, the number of all recorded codes and errors was considerably large to classify each of them into detect or not detect. Therefore, a 5% random sample of all records was classified. Consequently, we obtained 108 impasses to detect and 91 correct detections by the ruleset from the sample; hence, the recall was .843.

We reviewed the logs of misdetected impasses and found overdetections frequently. Moreover, we found misdetection because at condition (2-1) a trivial syntax error occurred while a student was resolving another compilation error; hence, the total number of errors did not decrease. To improve the detection ability, we adjusted the threshold parameters of the ruleset and modified condition (2-1) in Table 2. Table 3 provides our modified ruleset for impasse detection; the revisions are underlined.

Table 3: Modified ruleset for automatic impasse detection.

<table>
<thead>
<tr>
<th>Type of impasse</th>
<th>Conditions to detect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>(1-1) The learner has not modified the code for ( \geq 15 ) minutes.</td>
</tr>
<tr>
<td>Type 2</td>
<td>(2-1) The learner has not resolved identical compilation errors during four sequential compilations. (2-2) Compilation errors once resolved have reoccurred.</td>
</tr>
<tr>
<td>Type 3</td>
<td>(3-1) Same run-time errors have occurred ( \geq 7 ) times, successively.</td>
</tr>
<tr>
<td>Type 4</td>
<td>(4-1) Same position in the code has been modified by the learner ( \geq 4 ) times, successively. (4-2) Statement calling a standard output function, such as printf(), with some variable as arguments has been added and then removed. (4-3) Compilation has been executed ( \geq 5 ) times in ( \leq 2 ) minutes. (4-4) Code has been re-modified to a previous code. (4-5) ( \geq 25% ) of the code has been modified at once. (4-6) Statements including the same variable have been modified ( \geq 5 ) times, successively.</td>
</tr>
</tbody>
</table>

With the modified ruleset, 1836 impasses were detected from the same codes and errors. Of the total detections, 922 were correct, 756 were incorrect, and 158 were non-identifiable; hence, the precision was .549. We obtained 89 correct detections by the modified ruleset from the same sample containing 108 impasses to detect; hence, the recall was .824. \( F \)-measure, the harmonic mean of recall and precision, improved from .614 for the initial ruleset to .659 for the modified ruleset.
4. Overview of Our System Architecture

We developed the system visualizing the impasse summary, incorporating the subsystem for automatic impasse detection mentioned in the previous section. Java was the target programming language. Figure 1 displays our system architecture.

![Figure 1. Overview of our system architecture.](image)

Our system consists of the coding information tracker, the automatic impasse detector, and the impasse summary visualizer. The coding information tracker records for each learner every one-minute program-code that the learner is coding, every compiled program-code, every (if any) compilation error, and every (if any) run-time error. The automatic impasse detector scans the coding information in real-time, detects each learner’s impasse based on the ruleset described in the previous section, and records the type of impasse with coding information as the impasse log. The impasse summary visualizer summarizes the recorded impasse log and visualizes based on the user’s demand, namely, for reflective learning, one-on-one instruction during the exercise, or designing classroom. We have described the implementation of the coding information tracker and the automatic impasse detector as the subsystem in the previous section. Up to the present time, the impasse summary visualizer was also implemented only for learners’ reflective learning.

In reflective learning, learners reflected on the status of their understanding and identified the learning target that they had failed to understand. Hence, the impasse summary visualizer for reflective learning was required to visualize an overview of the types of impasses the learners had reached in exercises, and the snapshots of program-codes when they were experiencing impasses. Figure 2 provides the visualization of our impasse summary visualizer, including the snapshots of codes in (A), heat-map style summary of impasse detection in (B), impasse detection for each condition of the ruleset in (C), and interface of selecting a snapshot of the learners’ code in (D).

If all of the detected impasses were listed for the reflective learners, they would not understand what they should begin reflecting on. The visualizer provides the time-based summary of detected impasses in heat map style in (B), and each time slot is colored more deeply according to the more number of satisfied impasse conditions. Moreover, each satisfied condition is visualized on the time base in (C), presenting a period of detecting time. According to the visualizations in (B) and (C), learners clicked on a condition of impasse to reflect in (C), and then click on the system visualize buttons in (D) of the time slot included in the time period to detect the corresponding condition. When the learners clicked on a time slot button in (D), the system visualized the source codes at the beginning and end of the selected time slot.
5. Pilot Experiment

We conducted a pilot experiment to evaluate the contribution of our system to reflective learning. The experimental hypotheses are summarized as follows:

H1: Our system allows learners to easily perform reflective learning.
H2: The impasses detected and visualized by our system is useful for reflective learning.
H3: Our system promotes deeper reflection in reflective learning.

The subjects were five first grade undergraduate students majoring in computer science, who had participated in classrooms that had introduced our subsystem, experimentally described in Section 3. The targets of reflective learning were the actual exercises in the classrooms.

First, we explained the meaning and purpose of this experimental reflective learning and how to externalize the achievement of the reflection to the subjects in five minutes. Thereafter, we had the subjects perform the following three stages of reflective learning:

Reflection 1: Reflection depending on their memory and exercise materials only.
Reflection 2: Reflection with their impasse log recorded by our subsystem’s tracker and detector.
Reflection 3: Reflection using our system that includes a tracker, detector, and visualizer.
The subjects provided written briefs about their learning achievements at the end of every stage. After all of the reflective learning, we administered a questionnaire survey related to the subjects’ activities of reflective learning and the contribution of our system to it. The questionnaire contained the following five items on a five-point scale and a comment item:

Q1. How easy was it to perform reflective learning in Reflection 1?
Q2. How easy was it to perform reflective learning in Reflection 2?
Q3. How easy was it to perform reflective learning in Reflection 3?
Q4. How much did the impasses detected by our system contribute to reflective learning?
Q5. How instinctively correct were the impasses detected by our system?

Table 4 provides the responses of each subject for each questionnaire item. For each item, a higher score indicated a more positive response. The responses to Q1, Q2, and Q3 suggest that the subjects preferred reflection with impasse log to reflection without any support of our system, and preferred reflection with system’s visualization to reflection with raw log data. In addition, subjects #3 and #5 were exceptions; subject #3 preferred reflecting with raw impasse log because his coding time was extremely shorter than all other subjects and his impasse log was so small that he could easily observe his entire codes and errors without any summarization. However, subject #5 stated the most positive points for all reflections in his responses because his coding time was extremely long. He experienced impasses so frequently that all reflections sufficiently supported his understandings. On the other hand, his response in the comment item was that the easiest reflective learning was in Reflection 3. Based on this discussion, we consider that these results support H1. Moreover, the responses to Q4 and Q5 indicate that the subjects positively assessed the validity of impasse detections and the contributions of the detected impasses. We consider that these results support H2.

<table>
<thead>
<tr>
<th>Item</th>
<th>Subj #1</th>
<th>Subj #2</th>
<th>Subj #3</th>
<th>Subj #4</th>
<th>Subj #5</th>
<th>Ave.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3.00</td>
</tr>
<tr>
<td>Q2</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3.40</td>
</tr>
<tr>
<td>Q3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4.60</td>
</tr>
<tr>
<td>Q4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4.60</td>
</tr>
<tr>
<td>Q5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4.40</td>
</tr>
</tbody>
</table>

To verify H3, we reviewed the subjects’ briefings after each reflection stage. The details of the descriptions increased in the order of Reflection 1, 2, and 3, as a whole. Subject #1 described an impasse about Java methods at Reflection 1, while he described an impasse about methodization at Reflection 3. Subject #2 described the lack of understanding arrays at Reflection 1, while he described his concrete mistakes on processing Java arrays at Reflection 3. The other subjects also displayed similar behavior, describing details at Reflection 3, such as concrete mistakes and times in impasses, that were not found in the descriptions at Reflection 1. We consider that these evaluation results support H3.

We must consider that a small number of subjects may influence the accuracy of these verifications. Our discussions do not have sufficient reliability because we could not procure a sufficient number of subjects for the experiment. However, we believe continuous practice will suppress this matter. Based on these discussions, we conclude that although preliminarily, our system could support learners to perform reflective learning as a whole.

6. Conclusion

In this paper, we described automatic, rule-based impasse detection and the system supporting learners to perform reflective learning by visualizing the summary of the detected impasses. Learners’ impasse logs, which included the type of the impasse, the snapshot of program-code and the compilation and run-time errors at that time, would be useful for the learning and teaching of programming. Unlike learners coding effortlessly, those reaching an impasse are expected to experience some indication of
the impasse during their coding process. Therefore, we developed a ruleset to detect impasses based on the indications in learners’ codes and errors, and improved the detection ability by adjusting rules and threshold parameters through an experimental introduction into actual classrooms. Using this framework, we developed the system supporting learners to reflect how they had reached the impasse and how they had resolved it using the detected impasses as a clue. We conducted a small pilot experiment to evaluate the contribution of our system to reflective learning. The evaluation was based on a questionnaire survey and the achievements of reflective learning externalized by the subjects. The positive results of the evaluation suggest that our system could help learners attain some progress in reflective learning.

We must consider that a statistically insufficient number of the subjects in the experiment may influence the accuracy of verification. However, we believe that a continual practice of using our system will suppress this matter. Currently, our actual classrooms do not provide learners the time to reflect on their learning. We plan on introducing our system into the classrooms to make students cultivate a better understanding of programming exercises by performing reflective learning using our system. Persistent evaluations of students’ understandings would clarify the contribution of our system, and hence suppress this matter.

Currently, the impasse summary visualizer is implemented only for learners’ reflective learning. In the future, we will continue the development of the visualizer for teachers’ one-on-one instruction during the exercise and for classroom designing. We believe that teachers will not need only an overview of the impasse detections, but an overview of the correspondences between the learning target and each detected impasse. We plan to extend our system to address this issue, incorporating existing program evaluation systems. The existing systems (Konishi, Suzuki, & Itoh, 2000) evaluate the correspondences between learners’ code and sample code prepared by the teacher. Using this framework, we consider that our system could derive the correspondences between the learning target and detected impasse by integrating the learning target into each fragment of the sample code.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Numbers JP24300282, JP16K01084.

References


A Bibliometric Analysis of 15 Years of Research on Open Educational Resources

Xiaochen WANG\textsuperscript{a,b}, Mengrong LIU\textsuperscript{b}, Qianhui LI\textsuperscript{a} & Yuan GAO\textsuperscript{b}

\textsuperscript{a} Department of Educational Technology, Capital Normal University, China
\textsuperscript{b} Smart Learning Institute, Beijing Normal University, China
\* wangxc_cnu@163.com

Abstract: Open educational resources (OER) has developed for fifteen years since the term was adopted for the first time in 2002. To explore its research development progress as well as research focuses, this study reviewed literature on OER from the year of 2002 to 2017 with a bibliometric method based on seven indicators including publication year, distribution by country/territory, distribution by institution, journals, authors, essays and keywords. The findings include: 1. the top five productive countries of OER related research consist of Spain, USA, England, Romania, and China, 2. geographic proximity is a predominant factor in co-authorships, 3. International Review of Research in Open and Distance Learning is the most influential journal in the field of OER, 4. UK Open University is the most productive institution, 5. the research hot spots include OER, e-learning, higher education, mooc, open access, teacher training, ICT and open textbooks, innovation, pedagogy, web 2.0, etc., 6. there are three stages of OER research: emergence stage (2002-2007), exploration stage (2008-2011) and application stage (2012-2017), 7. policy and funding are two crucial factors influencing OER movement, and 8. sustainability, copyright, higher education, MOOCs are worthy of researchers’ attention. Hopefully this essay could draw a full picture of existing OER research, indicate its progress and frontiers, as well as provide implications for future work.

Keywords: Open educational resources, bibliometric analysis, CiteSpace, VOSviewer

1. Introduction

It is widely acknowledged that rapid development of information technology on one hand challenges traditional ways of education, and on the other hand, it provides new opportunities for teaching and learning (OECD, 2007). More specifically, “anyone can now learn anything from anyone at any time” (Bonk, 2009, p.6). In 2001, the MIT set up OpenCourseWare initiative (OCW), which made most MIT’s course contents available online for the public, and it could be regarded as the origin of now called “open educational resource movement”. The term “open educational resource” was adopted at a forum convened by United Nations Educational, Scientific and Cultural Organization (UNESCO) in 2002, and the latest definition is “any type of educational materials in the public domain, or released with an open license that allows users to legally and freely use, copy, adapt, combine and share” (UNESCO, 2017).

In the past 15 years, a continuously increasing number of institutions, researchers, teachers, and learners have taken interests and made contributions in research and practice of OER, but few attempted to focus on drawing a global picture of OER literature so as to present a complete development progress of OER research as well as relevant situations. From this perspective, the present study aims to fill up this gap by reviewing and analyzing existing research on OER. In other words, it tries to quantify and visualize academic performance and cooperation at multiple indicators including country, institution, journal and author, generalize development trend of OER studies, explore hot spots and frontiers, and provide implications for further research.

2. Methods
2.1 Methods of Data Collection

The data was extracted from the Web of Science (WOS) database, which is an authoritative and widely recognized database for science analysis. As a part of WOS database, Web of Science™ Core Collection database provides access to the world’s leading scholarly literature in the sciences, social sciences, arts and humanities and examine proceedings of international conferences, seminars, workshops, conventions, etc.

Data collection was conducted on May 11th, 2017. As the term “open educational resources” came into use from 2002, we choose that year as the starting point of literature review. We carried out the data collection within the Web of Science™ Core Collection database through the following steps. Firstly, check “advanced search”, “all languages”, and “all document types” in the option box, selecting timespan from 2002 to 2017. Secondly, search with topic keywords “open educational resources” and “oer”, which generated two collections of data with 2481 and 1786 records respectively. Thirdly, combine these two data collections with logical operator “OR”, which ensures that all relevant literature is included. Fourthly, refine data by the “EDUCATION EDUCATIONAL RESEARCH” category, and get a total of 910 records. Fifthly, export the records (including full records and references) to a txt document. The operation record is presented in Figure 1.

![Figure 1. Operation record of data collection](image)

2.2 Methods of Data Analysis

Bibliometric analysis is the main method of data analysis in this study, which is a quantitative methodology of published academic literature (Broadus, 1987; Mayr & Scharnhorst, 2015) and has been proved feasible in social science (Nederhof, 2006). According to van Raan (2003), bibliometric analysis could be performed at three levels: macro-level (e.g., countries), meso-level (e.g., institutions, universities) and micro-level (e.g., departments, research groups). His idea was adopted in terms of performance analysis.

With advanced tools, the exploration and visualization of distribution of publications, academic collaboration, development trend, and research focus can be effective and efficient. In this study, two software programs were adopted: CiteSpace and VOSviewer. Both of them are featured by visualization of literature. The former specializes in detecting and visualizing development trends and transient patterns (Chen, 2006), while the latter have superiority of clustering, network and density visualization. A combination of these two programs contributes to a more comprehensive explanation of OER literature.

3. Results

3.1 Publication Year

Figure 2 shows the publication year of OER literature, by which it can be seen that from 2002 to 2009, OER publications experienced a slow but relatively steady increase from 3 to 21. Then the figure increased by almost 3 times in 2010 to 66. After that, from 2011 to 2014, there was a stable growth from 66 to 134. In 2015, the figure reached the peak with a total of 182 publications. Although the number dropped a little in 2016, it still indicated a substantial outcome.
3.2 Distribution by Country/Territory

The publications in the past 15 years were distributed across 86 countries and territories, while the top 10 countries contribute to over 70% of the whole publications. As the Table 1 shows, Spain is the most productive country with 132 publications, which took up 14.5% of 910 records. The following four top productive countries are USA, England, Romania, and People’s Republic of China, respectively.

Table 1: Distribution by country/territory

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Number of Publications (NP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spain</td>
<td>132</td>
</tr>
<tr>
<td>2</td>
<td>USA</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>England</td>
<td>118</td>
</tr>
<tr>
<td>4</td>
<td>Romania</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>People’s Republic of China</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>Canada</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>Mexico</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>Germany</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>Australia</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>Italy</td>
<td>26</td>
</tr>
</tbody>
</table>

Figure 3 visualized the co-authorship network by country. Circle colors represent cooperation clusters. In other words, countries belonging to one cluster cooperated more with each other than with countries in other clusters. The size of circles refers to the number of publications of the country whose name is on it, and the lines stand for co-authorship relationship. Thicker a line is, more co-authorships existed between the two countries linked by it. This figure indicates that although there were a few long-distance co-authorships such as one between Canada and India, most co-authorships rely on geographic proximity. In addition, Spain, Germany, England, Italy and Greece are the top five countries with the highest co-authorships.
3.3 Journals

Number of publications and cited frequency are two important indicators of journal influence. Table 2 and 3 list the top 9 journals with the highest publications and citation frequency (of essays regarding OER) respectively. With the highest publications and citations, International Review of Research in Open and Distance Learning was obviously the most influential journal in OER research. British Journal of Educational Technology and Distance Education provide valuable insights as well.

Table 2: Journals with the highest publications on OER

<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal Name</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>International Review of Research in Open and Distance Learning</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>Open Praxis</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>EDULEARN14: 6th Annual International Conference on Education and New Learning</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>EDULEARN 11: 3rd International Conference on Education and New Learning</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>International Review of Research in Open and Distributed Learning</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>EDULEARN15: 7th International Conference on Education and New Learning</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>INTED2014: 8th International Technology, Education and Development</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>British Journal of Educational Technology</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>Distance Education</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 3: Journals with the highest citation frequencies of OER publications

<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal Name</th>
<th>Citation Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>International Review of Research in Open and Distance Learning</td>
<td>164</td>
</tr>
<tr>
<td>2</td>
<td>Computers &amp; Education</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>Journal of Interactive Media in Education</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>Open Learning: The Journal of Open, Distance and e-Learning</td>
<td>69</td>
</tr>
<tr>
<td>5</td>
<td>British Journal of Educational Technology</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>Distance Education</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>Interdisciplinary Journal of Knowledge and Learning Objects</td>
<td>57</td>
</tr>
<tr>
<td>8</td>
<td>Educational Technology &amp; Society</td>
<td>54</td>
</tr>
<tr>
<td>9</td>
<td>Educational Technology Research and Development</td>
<td>49</td>
</tr>
</tbody>
</table>
3.4 Distribution by Institution

Authors come from over 800 institutions, and Table 4 lists the top 11 productive institutions and their countries. UK Open University ranked as top one with absolute superiority over the second productive institution, Brigham Young University in USA, followed by Tecnológico de Monterrey in Mexico, Athabasca University in Canada and University of Nottingham in England.

Table 4: Distribution by institution

<table>
<thead>
<tr>
<th>Rank</th>
<th>Institutions</th>
<th>Country</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Open University</td>
<td>England</td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td>Brigham Young University</td>
<td>USA</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Tecnológico de Monterrey</td>
<td>Mexico</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Athabasca University</td>
<td>Canada</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>University of Nottingham</td>
<td>England</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>University of Alicante</td>
<td>Spain</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>West University of Timişoara</td>
<td>Romania</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>Utah State University</td>
<td>USA</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>Politehnica University of Timisoara</td>
<td>Romania</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>Universidad Nacional de Educación a Distancia</td>
<td>Spain</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>University of Bucharest</td>
<td>Romania</td>
<td>8</td>
</tr>
</tbody>
</table>

3.5 Authors

There are over 2000 authors including both first and co-authors. Table 5 shows the top 11 authors ranked by number of publications and citations. Wiley D, Holotescu C were the top two authors with ten publications per person, and Wiley’s publications have won the highest citation frequency. Besides essays, reports from some influential organizations like UNESCO and OECD were highly cited as well.

Table 5: Authors rank

<table>
<thead>
<tr>
<th>Rank</th>
<th>author</th>
<th>NP</th>
<th>author</th>
<th>Citation Frequency (CF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wiley D</td>
<td>10</td>
<td>Wiley D</td>
<td>111</td>
</tr>
<tr>
<td>2</td>
<td>Holotescu C</td>
<td>10</td>
<td>UNESCO</td>
<td>104</td>
</tr>
<tr>
<td>3</td>
<td>Grosseck G</td>
<td>8</td>
<td>Downes S</td>
<td>91</td>
</tr>
<tr>
<td>4</td>
<td>Hilton J</td>
<td>8</td>
<td>OECD</td>
<td>69</td>
</tr>
<tr>
<td>5</td>
<td>Mcandrew P</td>
<td>6</td>
<td>Atkins D E</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>Lane A</td>
<td>6</td>
<td>Siemens G</td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>McGreal R</td>
<td>6</td>
<td>Hylen J,</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>Stapleton S</td>
<td>6</td>
<td>Conole G</td>
<td>43</td>
</tr>
<tr>
<td>9</td>
<td>Andone D</td>
<td>5</td>
<td>Weller M</td>
<td>35</td>
</tr>
<tr>
<td>10</td>
<td>Montoya MSR</td>
<td>5</td>
<td>Mcgreal R</td>
<td>31</td>
</tr>
<tr>
<td>11</td>
<td>Beggan A</td>
<td>5</td>
<td>Dantoni S</td>
<td>29</td>
</tr>
</tbody>
</table>

3.6 Essays

Highly cited essays serve as essential knowledge resource and foundation in a research field, which can reflect development level, hotspots and advancing directions (Li & Zhang, 2016). Table 6 lists the top 15 cited essays. After reviewing essays in the list, we found that one third of the essays focus on the OER itself, trying to provide a full picture of OER. The detailed explanations of these essays will be discussed in the next section concerning hot spots.
Table 6: Top 15 cited essays

<table>
<thead>
<tr>
<th>Rank</th>
<th>Authors</th>
<th>Year</th>
<th>Essay</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Atkins D, E</td>
<td>2007</td>
<td>A review of the open educational resources (OER) movement: Achievements, challenges, and new opportunities</td>
<td>54</td>
</tr>
<tr>
<td>2</td>
<td>OECD</td>
<td>2007</td>
<td>Giving knowledge for free: The emergence of open educational resources</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>Downes S</td>
<td>2007</td>
<td>Models for sustainable open educational resources</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>Butcher N</td>
<td>2011</td>
<td>A basic guide to open education resources (OER)</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Weller M</td>
<td>2014</td>
<td>Battle for open: How openness won and why it doesn’t feel like victory</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>Hilton J</td>
<td>2010</td>
<td>The four ‘R’s of openness and ALMS analysis: frameworks for open educational resources</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>Yuan L</td>
<td>2013</td>
<td>MOOCs and open education: Implications for higher education</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>Geser G</td>
<td>2007</td>
<td>Open Educational Practices and Resources - OLCOS Roadmap 2012</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>Allen IE</td>
<td>2014</td>
<td>Opening the curriculum: Open educational resources in US Higher Education</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>Willems J</td>
<td>2012</td>
<td>Equity considerations for open educational resources in the globalization of education</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>Mcauley</td>
<td>2010</td>
<td>The MOOC model for digital practice</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>Liyanaguna wardena TR</td>
<td>2013</td>
<td>MOOCs: A systematic study of the published literature 2008-2012</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>Peter S</td>
<td>2013</td>
<td>On the role of openness in education: A historical reconstruction</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>Wiley D</td>
<td>2007</td>
<td>On the sustainability of open educational resource initiatives in higher education</td>
<td>5</td>
</tr>
</tbody>
</table>

3.7 Keywords

Table 7: Keywords list

<table>
<thead>
<tr>
<th>Rank</th>
<th>Keywords</th>
<th>Frequency</th>
<th>Rank</th>
<th>Keywords</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>open educational resources</td>
<td>175</td>
<td>12</td>
<td>ICT</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>oer</td>
<td>96</td>
<td>13</td>
<td>open textbooks</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>e-learning</td>
<td>64</td>
<td>14</td>
<td>distance education</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>higher education</td>
<td>58</td>
<td>15</td>
<td>innovation</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>mooc</td>
<td>44</td>
<td>16</td>
<td>learning objects</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>education</td>
<td>35</td>
<td>17</td>
<td>learning</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>open education</td>
<td>34</td>
<td>18</td>
<td>pedagogy</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>MOOCs</td>
<td>27</td>
<td>19</td>
<td>web 2.0</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>Online learning</td>
<td>20</td>
<td>20</td>
<td>educational technology</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>open access</td>
<td>18</td>
<td>21</td>
<td>blended learning</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>Teacher training</td>
<td>17</td>
<td>22</td>
<td>open educational resources(oer)</td>
<td>13</td>
</tr>
</tbody>
</table>

There are over 2000 keywords in the collection and Table 7 shows the top 22 appeared keywords. As the table shows, apart from OER itself, e-learning, higher education and mooc are mostly discussed.
Open access, teacher training, ICT and open textbooks, innovation, pedagogy, web 2.0, etc. are also hot issues related to OER.

To track the path of OER research focus, an analysis of keywords co-occurrence on timeline was conducted and the result was shown in Figure 4. There are three obvious clusters of keyword occurrence with representative words of “higher education”, “student”; “open educational resource”, “e-learning”, “educational technology”, “web 2.0”; and “learning object”, “mooc”, “innovation”, “online learning”.

4. Discussion

4.1 Development Trend

The publication year of OER research and time zone chart of keywords provide clues for development trend of OER research. As Figure 4 shows, there are three clear clusters of OER focus. With consideration of important conferences and forums on OER together, we concluded the development trend of OER research into three stages.

Emergence (2002-2007): In the early five years from adoption of term “OER”, there was a slowly increasing awareness of OER practice and study. The top keywords in this period include “higher education”, “education”, “resource”, “model”, “student”, etc. The main institutions involved in OER movement are universities, and students are at the center. Most studies concentrate on basic conceptual issues of OER, such as definition, characteristics and contents.

Exploration (2008-2011): In this stage, an overwhelming trend, web 2.0, has influenced education and OER significantly. Besides “open educational resource”, the top keywords in this stage also include “e-learning”, “educational technology”, and “web 2.0”, indicating that researchers tried to figure out technological issue of OER. In addition, international conferences in this period focused on awareness increasing and popularization of OER with the key interests of exploration of benefits that OER could bring to the public.

Application (2012-2017): A richer cluster of keywords appeared in this stage such as “mooc”, “technology”, “ICT”, “online learning”, “knowledge”, “innovation”, “learning object”, “teacher”, “connectivism”, “pedagogy”, “mobile learning”, “teacher training”, “learning analytics”, “open access”, etc. It indicates that OER research has reached a more comprehensive level; meanwhile, OER practice
has spread to broader fields such as innovation promotion, teacher training, and learning analytics. In accordance with keywords result, conferences during this period also focus on application of OER, including OER guidelines and policies.

4.2 Bursts Detection

The burst detection reveals sudden improvement of citation in countries, and the results were shown in Figure 5. Due to space consideration, we take England and China as examples for discussion.

<table>
<thead>
<tr>
<th>Countries</th>
<th>Year</th>
<th>Strength</th>
<th>Begin</th>
<th>End</th>
<th>2002 - 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>2002</td>
<td>14.0845</td>
<td>2005</td>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>ENGLAND</td>
<td>2002</td>
<td>10.6509</td>
<td>2009</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>GREECE</td>
<td>2002</td>
<td>3.2697</td>
<td>2013</td>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>PEOPLE'S R CHINA</td>
<td>2002</td>
<td>2.9183</td>
<td>2015</td>
<td>2017</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Citation burst detection

There is a solid evidence for the burst from 2009 to 2012 in England. It was in accordance with the duration of a national programme called UK Open Educational Resources (UKOER) programme. In this programme, over 80 projects have received funding and worked under the programme framework. The institutions involved has produced a great deal of academic outcome and experienced increased academic reputation. What is more, the projects not only focused on resource development, integration and popularization, but also attempted to address relative issues related to OER application such as the use of Creative Commons (CC) licenses in education, benefits gained by stakeholders, appropriate interfaces, successful business models, etc.

As for the case of People’s Republic of China, to a great extent, the burst appeared owing to an educational policy released by the Ministry of Education in late 2014, named <Implementation of effective mechanism of expanding high-quality educational resource coverage with information technologies>. Under this framework, it highlighted the “three accesses and two platforms” which stands for “every school has access to broadband networks, every class has access to qualified resources, every learner has access to online learning space” and two public service platforms for educational resources and educational management. The policy has stimulated investment and inspiration in OER development and research. Before long, in 2015, the Ministry of Education published an educational plan named <Strategic plan for higher professional education innovation development (2015-2018)>-. The policy pointed out the importance of developing co-construction and sharing system of digital resources, promoting development of high-quality and high professional resources based on market and regional demand.

The evidences shed light on the vital role of funding and policy in OER development and sustainability.

4.3 Research Focuses

According to keywords analysis, the research focuses of OER include e-learning, mooc, education, open education, online learning, open access, teacher training, ICT, open textbooks, distance education, innovation, learning objects, pedagogy, web 2.0, educational technology, blended learning, connectivism, evaluation, mobile learning, technology, distance learning, professional development, sustainability. With referencing to highly cited essays, we conclude four major topics concerned by OER stakeholders and researchers.

Sustainability is a core issue and major challenge faced by not only end-users but also OER developers, foundations and policy makers. Sustainability, in this context, refers to the ongoing ability of an OER project to make its goals come true (Wiley, 2007). Many may consider this issue from an
economic perspective and try to find an operational business model to sustain OER projects in long-run. However, sustainability is not restricted to financial problems, and it is unrealistic to build a one-fit-all model owing to the differences of OER projects (Wiley, 2007). Although funding is important, in a larger picture, calculation of other surrounding issues such as content management, organization operation and technical maintenance, is also necessary (OECD, 2007). Therefore, Downes (2007) categorized the sustainable OER models into four aspects: funding, technical, content and staffing. Wiley (2007), argues for two main challenges for OER projects’ sustainability: production and sharing, as well as use and reuse.

Copyright issues are at the heart of OER (Atkins, Brown & Hammond, 2007). There is an inevitable confliction between rights-holders and public good, so the thing need to be done is seeking for an appropriate balance, which is also the basic rationale of open licenses. But there is a common misconception: content released under an open license belongs to the public, so all users can adapt and reuse it (Butcher, 2011). This could be verified by the OER evidence report (de los Arcos, Farrow, Perryman, Pitt& Weller, 2014), which found that over 80% of informal learners make adaption to resources, but only 18% of them do it after confirmation of their rights to do so. In fact, none of open licenses is a “yes” or “no” switch, there are different levels of openness. The most widely applied open license in OER, Creative Commons license, defines seven levels of openness (from the most openness to the least): CC 0, Attribution, Attribution-ShareAlike, Attribution-NoDerivates, Attribution-NonCommercial, Attribution-NonCommercial-ShareAlike, Attribution-NonCommercial-NoDerivates. Under this framework, right holders of resources can decide their rights to reserve. At the same time, the public can use these resources within the permissive space.

Higher education is the main sector for OER movement. This is understandable given the fact that in the information age, all countries are trying their best to ensure citizens have access to quality tertiary education, but universities do not open their gates to everyone, especially the “elite” ones. That is why the OpenCourseWare initiative broke the wall of university and provide high-quality educational resources to the public. After that, an extraordinary sharing trend spread through worldwide institutions, organizations and even individuals. They made great contributions to social equity, not nationally but globally. Although higher education may be the main filed of OER still, K12 education and professional training are making efforts to keep in pace.

Massive open online courses (MOOCs) was originally used to describe a course called “Connectivism and connective knowledge”, which was initially designed for 25 formal students to study for credit and open registration of worldwide learners, and surprisingly got over 2300 registers by its end (Yuan &Powell, 2013). The emergence and development follows the trend of openness in education and OER movement (Yuan, MacNeill & Kraan, 2008). Some may have confusion between OER and MOOCs. In fact, MOOCs, as complete courses including instructions and assessments, could be regarded as a subset of OER, which consist of learning content, tools and implementation resources. However, a great deal of open educational resources, appeared as the formation of MOOCs such as Coursera, Udacity and edX.

5. Conclusion

It is undeniable there are limitations of this study. Firstly, there is no possibility that a database can cover all the publications in a fast growing filed (Wang & Liu, 2014), not to mention that research outcomes are published in various languages. Therefore, it is inevitable that we might miss some articles due to the limitation of database and data collection. Nevertheless, Web of Science database provides a relatively comprehensive collection of core academic outcomes, which help us to obtain a relatively rational results. Secondly, due to space consideration, some findings could not be discussed in detail. Given the fact that the goal of this essay is to provide a full picture of OER research rather than a deep analysis, we strongly suggest reading essays in terms of specific issues which can provide deeper insights.

To sum up, this study reviewed OER literature from 2002 to 2017 with a bibliometric method. The results include publication years, distribution by country, institution and journal. Influential authors and essays were also extracted. With the results related to publication year and keywords clusters, the development of OER research falls into three stages: First stage of an emergence, during which the
researches focus on the definition, features and contents of OER, and tried to increase the public’s awareness. Second stage of an exploration, during which the research focus referred to technological issues and serviceability; third stage of an application, during which the research focus became more comprehensive and practical. The burst detection in countries indicate that policy and funding are two crucial factors influencing OER movement, consequently are much concerned by stakeholders. Apart from them, sustainability of initiatives, copyright issues, higher education, MOOCs are worthy of researchers’ attention. Although some issues encountered just once such as innovation and pedagogy, they should not be overlooked. Instead, both practice and research should be placed in a full picture and considered carefully.

Acknowledgements

We would like to thank funding and support from Research on Co-creation, Sharing, Promotion and Implementation of Open Educational Resources along the Belt and Road Initiative, A national general project in educational science in “the 13th Five-year Plan”. 2017, BCA170073.

References

Wiley, D. (2007). On the sustainability of open educational resource initiatives in higher education. Paper commissioned by the OECD's Centre for Educational Research and Innovation (CERI) for the project on Open Educational Resources.
Design and Implementation of a Pedagogic Intervention Using Writing Analytics

Antonette SHIBANI*, Simon KNIGHTa, Simon BUCKINGHAM SHUMa & Philippa RYANb

aConnected Intelligence Centre, University of Technology Sydney, Australia
bFaculty of Law, University of Technology Sydney, Australia
*antonette.aileen.shibani@student.uts.edu.au

Abstract: Academic writing is a key skill required for higher education students, which is often challenging to learn. A promising approach to help students develop this skill is the use of automated tools that provide formative feedback on writing. However, such tools are not widely adopted by students unless useful for their discipline-related writing, and embedded in the curriculum. This recognition motivates an increased emphasis in the field on aligning learning analytics applications with learning design, so that analytics-driven feedback is congruent with the pedagogy and assessment regime. This paper describes the design, implementation, and evaluation of a pedagogic intervention that was developed for law students to make use of an automated Academic Writing Analytics tool (AWA) for improving their academic writing. In exemplifying this pedagogically aligned learning analytic intervention, we describe the development of a learning analytics platform to support the pedagogic design, illustrating its potential through example analyses of data derived from the task.

Keywords: academic writing, writing analytics, learning analytics, learning design, pedagogy

1. Introduction

Academic writing is a key professional skill for students to develop. Despite its importance, students are seen to lack proficiency in writing (National Commission on Writing, 2003). Support for academic writing has been limited, mostly taking the form of English for Academic Purposes (EAP) for non-native speakers, or remedial action to improve writing skills in an ad-hoc manner (Wingate, 2012). There is a need to help students with their academic writing in an ongoing and integrated way. Formative feedback on writing aids students in gaining awareness regarding their progress against their goals. Through such formative feedback, students can close the feedback loop by applying the feedback that they receive to improve their work, to address the gap between their performance and instructor expectations. This approach arguably results in greater impact on students’ learning than summative assessments (Sadler, 1989).

However, for large classes, it is not practically possible for the instructor to provide formative feedback to all students since the process is time-consuming. To overcome this issue, automated tools have been developed that use computational techniques to assess writing. The scope of such tools varies from Automatic Essay Scoring (AES) systems that provide a score based on the assessment of standardized writing to Automated Writing Evaluation (AWE) systems that provide additional feedback to students on their writing (Warschauer & Grimes, 2008). Learning Analytics, which makes uses of analytics techniques on student data to improve learning, can be used for providing formative feedback which is almost immediate. Several tools have been employed for university and school students to analyse text in the context of essays, problem solving, free form and collaborative writing. One such tool is the Academic Writing Analytics (AWA) tool that provides formative feedback on students’ academic writing. AWA uses natural language processing techniques to identify sentences in a text that match specific rhetorical functions, like emphasizing an important point or summarizing, by using linguistic markers that indicate these rhetorical moves (Knight et al., 2016). These kinds of moves are a key component in good academic writing and are seen to be correlated to essay quality (Simsek et al., 2015). Feedback on the presence of these moves should help students reflect on their writing and the rhetorical structure of it.

Regardless of the quality of such technology, a concern in technology-enhanced learning is that technologies may not be used unless they are embedded in the curriculum (Wingate, 2012).
alignment of learning analytics to learning design has also been increasingly emphasised to provide a contextual framework for the pedagogic intent of analytics applications (Lockyer, Heathcote, & Dawson, 2013). A clearly defined pedagogical design closes the gap between the potential and the actual use of technologies, by helping students put these tools to appropriate use in order to add value to their learning. This forms the basis for learning analytics pedagogic interventions design which moves from developing learning analytics technologies to integrating them as part of a larger educational context (Wise, 2014). The integration of learning analytics tools in pedagogic design should also be aligned to subject curriculum in order to find new ways of solving existing pedagogical issues using learning analytics. Good design of learning analytics platforms also makes collection of data much easier, which can give useful insights for guiding students during the length of the course and in future interventions. The aim of this study was therefore to design an effective pedagogical intervention and a learning analytics platform to introduce the automated writing analytics tool AWA to students to help them write better essays for their subject. The contribution of this paper is to provide an exemplification of a pedagogically aligned learning analytics intervention and platform developed to gain research and learning insight into student writing.

2. Techno-Pedagogical Design: Aligning Learning Analytics with Learning Design

2.1 Research Context

This study was designed for 367 undergraduate law students enrolled in a Civil Law subject. The study described was co-designed with the instructor and piloted by students and subject tutors as a pedagogic intervention in one of their weekly tutorial sessions. Writing is a key disciplinary skill for law students with emphasis on clear and engaging writing with the use of appropriate arguments (Knight et al., 2016), and was identified as an area to target student learning. As part of their curriculum, law students are expected to write academic essays that discuss an assigned topic, clearly outlining the legal arguments. In their key written assignment, the instructor has developed a marking rubric consisting of the following elements: Statement of argument, Statement of essay plan, Identification of issues, Analysis, Sustained thesis & Original insight, and Engagement with literature/cases. The intervention designed required students to complete an online activity in class during a tutorial session consisting of several sub-tasks. The rationale was to help students write better essays during the course of their subject by understanding the instructor’s rubric for assessing an essay and practising revision skills.

2.2 Intervention and Platform Design

The objective of the study was to design a pedagogic intervention for students to improve their ability to evaluate the quality of writing and revisions and improve a draft based on feedback/self-assessment. The design was developed to provide both a learning experience for the students and to facilitate the development and evaluation of learning analytics interventions by the researchers. The sequence in Figure 1 is a simplified workflow of tasks that the students will carry out. All these tasks will be explained along with the pedagogical reasoning behind their design in this section. They will also be carrying accompanying technical information on how these tasks were implemented in a learning analytics platform.

Students were randomly pre-assigned to one of the three groups by the instructor based on the feedback they would receive in the revision task as follows:
1. AWA Feedback Group – These students received feedback on request from the AWA tool. They can request feedback on their revised text as many times as required. They watched a short video on how to use the tool at the start of the activity.

2. Instructor Feedback Group – Students from this group saw an instructor-highlighted PDF file with static feedback on the parts that need improvement, text that could be improved, and the text that is a good example of academic writing based on the given essay.

3. No Feedback Group – This set of students received no feedback on the essay to make improvements. They worked on the text based only on their assessment of it.

These conditions were designed to address a key concern of the research questions: understanding the efficacy of different feedback types for student revision, in order to identify helpful feedback for students.

Students worked individually on the activity using their own laptops. They entered the system by accessing the platform from the web. The URL for the activity was supplied from the LMS as part of their weekly lesson. The tasks were almost identical for students from all the groups, except for the feedback received to complete the revision task. Student details were pre-stored in the database and each student was directed to a specific group’s tasks upon login. The pedagogical design was supported by a learning analytics platform that facilitated the online activity with several sub tasks. The technical platform (developed in PHP) for the activity was built to be scalable and flexible in order to adapt to the needs of learners and instructors in different contexts. The platform architecture of how this activity was implemented is shown in Figure 2. Traces of student activity are stored in database tables by different components of the web interface as students use the platform. The time spent for the whole activity to complete all tasks from the start to the end is also recorded for all students. The different subtasks in the activity and student activity data stored by different components of the web interface are explained in the following sub-sections.

![Figure 2. Platform architecture](image)

### 2.2.1 Rubric Understanding

In our learning context, although students are already aware of the instructor’s marking rubric, they may not know how to apply the rubric, and how particular rubric facets are related to linguistic features that automated tools might help them identify. Therefore, the first task was a matching exercise (implemented using a customized DHTML drag and drop quiz script [http://www.dhtmlgoodies.com/index.html?whichScript=drag-drop-quiz](http://www.dhtmlgoodies.com/index.html?whichScript=drag-drop-quiz)), where students were asked to identify sample sentences from an essay that would match elements of the instructor’s marking rubric. This engagement with exemplars is seen as an effective method for students to understand the assessment criteria (Hendry, Armstrong, & Bromberger, 2012; Rust, Price, & O’donovan, 2003). The task thus supports understanding the different rhetorical markers from sentences that would be useful to signal to the reader the important components of their essay with respect to the rubric. In the AWA feedback group, students also saw the corresponding AWA tags for the exemplar sentences to support understanding how the tags were related to the rubric. A screenshot from the matching exercise is shown in Figure 3, where green indicates correctly matched elements and red indicates wrongly matched elements.
matched elements. Students were required to match all instances correctly before moving to the next task. In the analytics platform, the matching exercise stores the duration taken by students to complete this task by matching all elements correctly.

![Figure 3. Sample examples from the matching exercise](image)

### 2.2.3 Understanding Task Requirement

In the second task, students viewed a sample essay which was revised by the instructor to give them an idea of the kind of features to focus on and how revisions to an existing text might be used to improve that text towards the rubric. The process of revision could otherwise include anything from making surface level changes like spelling and grammar to modifying the content of the topic. For this particular revision task, students were encouraged to focus on rhetorical structures in the text that could be improved. This task did not collect any data for assessment, but was designed for students’ understanding only.

### 2.2.4 Essay Assessment

The third task consisted of an essay assessment in which a low quality essay exemplar was provided. Students first provided an assessment of the essay’s quality, which, in the fourth stage, they then worked on to revise. This task was designed for students to acquire evaluative expertise by transitioning from feedback to self-monitoring (Sadler, 1989), with the self-assessment intended to enhance students’ capacity to make judgements and self-regulate their work for sustainable learning (Boud, Lawson, & Thompson, 2015). The essay assessment component stores students’ assessment data on the given essay in the form of grades, confidence level that it would match an instructor’s grading, and qualitative comments on the problems identified in the essay and recommended improvements.

### 2.2.5 Essay Revision

Students thus, fourthly, worked to act on the issues that they identified in the text, with encouragement to engage in the kinds of revisions cycles that support learning to write (MacArthur, 2007). In this task, students received different types of feedback on the essay to make revisions based on their group (AWA Feedback Group, Instructor Feedback Group and No Feedback Group). To facilitate the use of feedback, it was provided in a frame to the right of the editor frame in which they revised their text. The revision task interface for Instructor Feedback Group is shown in Figure 4. For the AWA feedback group, the frame on the right contained feedback on the editor text from the AWA tool. A sample analytical feedback from AWA is shown in Figure 5. The no feedback group was provided with the text editor only to make revisions. A basic document editor from CKEditor was used for the revision task ([http://ckeditor.com](http://ckeditor.com)). This preserves formatting of text which would be lost in a normal text box. Text cleaning and formatting were then performed in PHP to provide live feedback on the text, or post-task processing. As mentioned before, students in the AWA group could get feedback from the tool on any
of their drafts as needed. The analytical engine that provides feedback on the text was accessed from a version of Xerox Incremental Parser (https://open.xerox.com/Services/XIPParser). After revising the essay, students completed a self-assessment on the revisions made, to reflect on the improvements they made in the essay.

For this revision task, students can be provided with different types of feedback as required by the pedagogical design. In the current setting, there was a static feedback by the instructor in the form of a pdf which gave comments on the given essay, and a dynamic feedback from the automated analysis tool AWA that can be accessed to get feedback at any particular version of the essay as required. This sends and receives data from an external parser to provide feedback on analytical writing. Data cleaning might be required to format the data from the editor in the correct format required by parser. From the provided text editor, data was also stored in specific intervals (every one minute) to capture students’ drafting process. The final improved essay was stored in the database for all students. The dotted database tables in Figure 2 show data that will be different for different students based on their usage behaviour. The number of draft essays stored depends on the time they spent working on the task—higher time spent stores more drafts. AWA use also varies from student to student – ranging from students who made few requests for feedback to students who requested feedback many times.

2.2.6 Feedback Survey

In the final part, students provide feedback on the task and the feedback obtained for revision by answering a few questions. At the end of the activity, students were provided with an option to download a version of their revised essay (dynamically generated using http://www.fpdf.org/) and a
sample revised essay by their instructor for future reference. This was to help them reflect on the improvements they made in the essay by comparing with an exemplar improved essay from the instructor. The feedback survey stores responses from students on the questions asked about the usefulness of task and feedback. This consists of rating responses and qualitative comments. Details of students who downloaded their own improved essay and instructor's sample revised essay from the last page for reflection are recorded in the database as download history by tracking the clicks from the respective links.

3. Data Analysis

Development of the learning analytics platform and the pedagogical design described above facilitates capturing student trace data that can be analysed to provide insights into their learning, and the impact of feedback on that learning. Student activity data stored in the database tables can be downloaded as csv files, which could be imported into R for data analysis. Even though the activity was carried out in a tutorial session in class, not all enrolled students completed the activity since it was not a mandatory requirement. For the purpose of this study, only the complete dataset of 201 students who finished all parts of the activity is considered for analysis. This consisted data of 91 students from AWA group, 71 from instructor group and 39 from no feedback (none) group. Preliminary data analysis below shows a sample analysis that can be performed from a subset of the available data.

3.1 Student Perceptions of the Usefulness of the Pedagogic Design

Students rated the perceived usefulness of this activity in order to improve their essay writing in a scale of 1-5 (where 1= not at all useful, 2= slightly useful, 3= somewhat useful, 4= very useful and 5= extremely useful). They further provided qualitative comments on what feedback was found to be useful, what feedback was not useful, and any other additional comments about the whole activity. This data was analysed to see students’ feedback on the perceived usefulness of the activity and on the provided feedback (instructor and automated tool feedback).

The perceived usefulness of the activity across the three groups is shown in Figure 6. The instructor feedback group found the activity to be most useful (M = 3.34, SD = 0.71), followed by no feedback group (M = 2.92, SD = 1.07) and AWA feedback group (M = 2.80, SD = 0.56).

![Figure 6. Perceived usefulness of the activity across comparison groups](image)

A one way analysis of variance showed that the effect of groups on the usefulness score was significant, F (2,198) = 8.32, p = 0.0003. The assumption of homogeneity of variance was satisfied with Levene’s test (p = 0.12). Post hoc analyses using Tukey’s HSD indicated that the usefulness rating of instructor feedback group students was higher than the usefulness rating of AWA group students (p=0.0002), and no significant difference was noted amongst the other groups.

The qualitative comments of students were explored to understand their views on this activity. The no feedback group provides a baseline group, as these students evaluated the usefulness of the
broad pedagogic activity with no additional feedback component. Across the groups, students found several sections of the task useful in improving their essay writing. A number of students explicitly mentioned that it was useful to have the initial sample text, on which the instructor had modelled the kinds of revisions that could be made to improve a draft, saying things like:

- “The annotated sample with comments was helpful in revising the essay, as it gave examples of what was done well and done poorly.” Respondent 123, Instructor feedback group
- “The exemplar answers were very helpful in highlighting the areas of the essay which needed improving, which students may initially overlook. The highlighting of different sentences is also useful in indicating what components of the writing were critical and what sections were maybe unnecessary identification or description. The sentences crossed out and rewritten were especially useful for proposing alternative ways of writing a sentence in an improved manner” Respondent 82, AWA feedback group
- “It was useful to see how a simple change (like swapping one paragraph for another) can make an essay a lot clearer and relevant to the topic at hand. I will make sure I apply this kind of task to my own essays - asking myself if my essay would be clear to a pair of fresh eyes.” Respondent 180, No feedback group

Students also appreciated having access to both their own text to download, and a sample revision of the same text they had edited which the instructor had marked up with improvements, saying for example:

- “A very good exercise! Glad we can download both our improved version and the instructor's improved version. Hopefully this will be a way we can get feedback on the feedback we provided in our edit.” Respondent 145, Instructor feedback group
- “The provision of the instructors improved essay provides a useful benchmarking tool to compare my changes against the changes made by the instructor” Respondent 109, Instructor feedback group

3.1.1 Evaluating Design Decisions

Some students felt that it would have been useful to have some readings beforehand on the topic to have a better idea on the essay they work with.

- “I really love the idea behind this exercise. I think it would be more beneficial to complete if we had to do some prior reading - for example of the Salyzyn essay that was referred to in the paper so that we could have some context. I find it difficult to write or revise something without having a background in the area.” Respondent 179, No feedback group
- “It didn't make sense to be asked to revise an essay on a topic we haven't really studied ... because in terms of content, I'm not sure how to improve it” Respondent 28, AWA feedback group

This could be incorporated if a revision task is designed in the future for modifying the content of the essay as well. The focus of the current task was on the rhetorical structure improvements and hence there was no emphasis on the content.

Some students were unsure of the usefulness of the revision task which required them to work on essays written by others. On the other hand, few others found it useful to apply their critical lens to an essay written by someone else, as it would eventually help them look critically at their own essays.

- “Everyone has their own unique styles that should translate on to a page of work that is of their own design. Not sure if people learn from someone else's mistakes at the very end of a Uni day.” Respondent 167, No feedback group
- “I believe having to personally assess an essay forces you to critically engage to a greater extent than one may have to. From looking at an essay from a marker's perspective one can take a step back and understand the little details that a marker is looking for. I also feel that by assessing someone else's work it provides you with better skills to assess your own work from a more neutral perspective. Self-reflection and editing are a key aspect of writing a poignant and quality academic essay that accurately engages with the criteria.” Respondent 32, No feedback group

The allocation of time was not incorporated in the platform, but was provided as a run sheet with an approximate time division for the sub-tasks. This led to some incomplete submissions as some students stayed in the first few tasks longer than expected, not allowing them enough time for the rest of the activity. The activity could be re-designed in the future to incorporate time allocation for sub-tasks in
the platform for smooth task completion. It could also be built as an out-of-class activity where students can engage with the tasks in their own time out of class, independent of the pace of other students.

- “The activity was engaging but it would be more enjoyable if it was clear of the time allocation for the tasks and the number of tasks involved” Respondent 204, No feedback
- “I think it would have been useful to know how much time I had...I feel as though I rushed myself with the editing and therefore didn't do my best work” Respondent 109, Instructor feedback group
- “A bit more time in amending the essay would have been greatly helpful.” Respondent 135, Instructor feedback group

3.1.2 Evaluating the Provision of Meaningful Feedback to Students

In terms of the feedback provided to improve their essay, students from the instructor feedback group felt that more explanations were necessary to understand the changes that they needed to make in their essays. They wanted to learn how to resolve the identified problem by receiving suggestions for improvement. Such direct recommendations would help students solve the current problems in hand and aid them in resolving similar problems in the future. Similar comments on the given feedback were also seen in the AWA group. They wished to receive more direct feedback in the form of corrective advice on what to improve, rather than highlighting the key sentences.

- “Rather than just highlighting the text, I think it would have been worthwhile to have colour coded explanations of why each section was highlighted… Sometimes I was unsure about why a piece of text was highlighted so I wasn't sure how to make a change.” Respondent 109, Instructor feedback group
- “Didn't give many alternatives as to how the phrasing could be improved” Respondent 143, Instructor feedback group
- “The highlighting only alerted to me what was good. However, there should be highlight to alert me to problems in the essay as well. the highlighting only showed me what was a 'summary' etc. There should be more categories and types of feedback such as grammar issues, sentence structure.” Respondent 11, AWA feedback group
- “It could offer an alternative or some tips regarding essay writing so a student who has seen where they go wrong can understand how to amend the essay they have written.” Respondent 95, AWA feedback group
- “I found the feedback unhelpful as I couldn't distinguish which parts needed fixing, even though it was stated as "important". Couldn't understand what 'important' meant in this context - important to fix or important as in it was a good part of the essay that didn't need to be fixed?” Respondent 47, AWA feedback group

Student feedback from this task could be used to find ways in which the tool can be tweaked to provide better feedback. However, few of the suggestions students made with regard to the feedback received provide direct actionable alterations for that feedback. It is crucial for students to recognise the intended usage of the tool and how to use it best to help in their writing context. The scope of the tool must be explained clearly in terms of what it can and cannot do. For example, some students noted that they would have liked grammar and spelling feedback, but this is not a feature targeted by the task or the tool deployed in this research. Students noting that the focus of the tool is on rhetorical markers of academic writing should set expectations of the tool to students in order to effectively use the tool. Further guidance regarding the use of the AWA tool, in the form of examples of use, and a user-guide, would also support this effective use. Directing students to lessons where they can read more about effective writing practices would also be beneficial.

Some students were not comfortable in receiving automated feedback and felt that a tool cannot provide context-sensitive feedback like a human. This is a known problem with the incorporation of such tools. Students should be made aware that automated tools are not a replacement for instructors/tutors, but rather a support mechanism they can use to get additional feedback when required. As discussed earlier, making students understand the scope of the tool would help them put the tool to appropriate use it was designed for. This will help them know the context of using a machine versus a human for the desired feedback.

- “I don't feel as comfortable with an online tool. I think I would feel more comfortable with a human providing feedback” Respondent 21, AWA feedback group
“An automated program would not be able to tell me what points I am missing information-wise like a lecturer or tutor would be able to” Respondent 25, AWA feedback group

Students who received automated feedback from AWA also rated their level of comfort (1-5, where 1 = not at all comfortable and 5 = extremely comfortable) in receiving feedback from a tool. Students were generally not comfortable in receiving feedback from a tool (M = 1.25, SD = 1.56). Students’ comfort level in receiving automated feedback was also found to be positively correlated to their perceived usefulness of the activity, Pearson’s r(199) = .44, p < 0.0001. This could provide a possible explanation for the low usefulness score of the AWA feedback group. The low usefulness score of AWA group students in this activity is in contrast to the findings from the previous study (Knight et al., 2016) which reported that students found AWA to provoke useful reflection regarding their essay writing. Students seemed to have judged the usefulness of the current activity in terms of their tool usage rather than the wider pedagogic design of the tasks targeting improving their essay writing. Thus, the expectations for different types of feedback by the tool and their comfort level to receive automated feedback could have contributed to the usefulness score for this group.

3.2 Further Analysis and Platform Capabilities

A limitation of the current design is that the intervention is run as a one-time study, so how this activity contributes to improved essay writing of students in the long term is not studied. A continuous assessment of the influence of this activity on students’ overall improvement in writing skills would be beneficial. This could be conducted by investigating future essays and their corresponding grades.

From this existing dataset, a detailed analysis is currently underway. The improved essays that students submitted are being graded by the tutors using the rubric. This additional data allows us to address questions around (1) the impact of different kinds of feedback on performance, at a criterion level; and (2) to associate different editing behaviours with performance, again at a criterion level. The former dataset provides an evaluation of feedback, while the latter provides further insight into the kinds of revisions we might target our feedback at. Specifically, the dataset provides for analysis of tool usage, time spent on the task, click stream history of downloads, demographics, and particular revisions made. The different types of revisions made by students could be characterised in terms of surface text metrics (how much text has been edited, word count, sentence length, cosine similarity to the original document) or with regard to particular features (including the introduction/deletion/editing of rhetorical features in the text, cohesion, etc.). These metrics can then be studied with respect to the marking grades, which can provide useful measures for automatically assessing text quality in the future.

In addition, the platform collected 1 minute (an adjustable parameter) snapshots of students’ writing. These can be analysed to study the drafting process of different students in terms of number of revision actions like addition, substitution, deletion etc. and how these contribute to the final improved essay. Given that students identified issues in the original text alongside an assessment of that text, we can also investigate the relationship between the student assessment (both grade and qualitative comments) and their revisions, particularly with regard to their subsequent self-assessment of the improved draft. The interpretation of results from this data can be presented to students and instructors as reports to provide feedback on their writing and teaching practice. There is thus a huge potential for the analysis of data collected to provide insights to researchers, as well as instructors and students.

4. Discussion and Conclusion

In this study, we designed a pedagogic intervention to support student learning, and both implemented and evaluated the potential of the writing analytics tool AWA. The pedagogic design and development of a learning analytics platform to support the intervention exemplifies a learning-oriented approach to learning analytics. Traces of student data made available by the platform enable many types of analysis with the use of quantitative, qualitative and data science techniques. These can be used to provide insights for both students and educators. Based on our preliminary findings, students generally found this activity useful in developing their writing. Qualitative analysis of students’ comments on the activity sheds light on the usefulness of different subtasks in the pedagogic design that contribute to their writing skills. It was observed that students found the exemplars, self-assessment and revision
skills applicable for their own writing in the future. Such interventions are seen to improve students’ understanding of the instructor’s rubric and their writing skills by deliberate practice of these skills.

Analysis of students’ comments regarding the feedback they received provides information to instructors and researchers on the types of feedback that students find useful, and their expectations of feedback from instructors and automated tools. Students particularly highlighted a desire to receive explanations on why certain sections are highlighted and how to improve the text further. Giving such actionable feedback for students to close the gap between the expected and current performance is a principle of good feedback practice, which has to be followed for any kind of formative feedback to be given to students (Nicol & Macfarlane-Dick, 2006). When tools are used, students also require proper guidance to interpret the results and use the tool for its intended purpose. These findings will help researchers to evaluate automated feedback, and thereby design tools that provide meaningful and actionable feedback to help students in their academic writing. It is also crucial to design effective pedagogical practices keeping in mind the inherent limitations of automated tools.

The pedagogic intervention could be re-designed in many ways to suit the educational context in hand across different scenarios. The same activity could be implemented for other subjects by modifying the instructor rubric elements and exemplars pertaining to that particular subject. The pedagogical elements of the activity can be modified by adding and removing the sub tasks as required. The task could be re-designed to evaluate the different types of feedback from other tools and other forms of feedback from an instructor. Different versions of a tool can also be tested by using a specific portion of the activity. In another setting, students could also use this tool to work on their own writing for assessment and improvement instead of an assigned writing. These changes can be easily made in the platform by modifying specific components. Thus an alignment between learning design and learning analytics fosters both pedagogically-grounded activity for students, informs the design of automated feedback, and generates research data for the evaluation and development of novel writing analytics tools.

The pedagogic intervention could be re-designed in many ways to suit the educational context in hand across different scenarios. The same activity could be implemented for other subjects by modifying the instructor rubric elements and exemplars pertaining to that particular subject. The pedagogical elements of the activity can be modified by adding and removing the sub tasks as required. The task could be re-designed to evaluate the different types of feedback from other tools and other forms of feedback from an instructor. Different versions of a tool can also be tested by using a specific portion of the activity. In another setting, students could also use this tool to work on their own writing for assessment and improvement instead of an assigned writing. These changes can be easily made in the platform by modifying specific components. Thus an alignment between learning design and learning analytics fosters both pedagogically-grounded activity for students, informs the design of automated feedback, and generates research data for the evaluation and development of novel writing analytics tools.

**References**


Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 203-211).*
Multimodal Interaction Aware Platform for Collaborative Learning

Aoi SUGIMOTO*, Yuki HAYASHI & Kazuhisa SETA
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan
*sugimoto@ksm.kis.osakafu-u.ac.jp

Abstract: A number of studies in the research field of Computer Supported Collaborative Learning (CSCL) have proposed various systems in order to facilitate learning in the context of social interactions. In a collaborative learning, participants exchange not only verbal but also nonverbal cues such as utterance, gaze and gesture for maintaining the relationships among one another. Nevertheless, very few attempts have been made to construct a CSCL system that can utilize such multimodal (verbal and non-verbal) information to support communication in a collaborative learning. In this paper, we propose a novel platform that enables CSCL system developers to construct their learning support tools that have original functions to process such multimodal information. By building learning support tools on our multimodal aware platform, we confirmed its usefulness and also potential to pioneer unexploited filed of research in learning analytics for CSCL and methods to intervene in collaborative learning processes using verbal and non-verbal information unutilized so far.

Keywords: CSCL, Multiparty Multimodal Interaction, System Development Platform, Verbal and Non-verbal Information

1. Introduction

In our globalized cross-border world, we are required to cultivate social interaction skills that enable us to collaboratively make decisions or solve problems with other colleagues in various situations (Griffin, et al., 2012). In order to cultivate such social interaction skills, meaningfulness of collaborative learning style based on social constructivism whereby plural participants acquire knowledge and solve a problem is widely recognized. For a successful collaborative learning whereby all the participants can get fruitful learning outcomes, participants’ mutual engagements to the learning processes are required in addition to solving the problem itself. Participants get several benefits from such interactions, going from constructing deeper level learning, shared understanding, to developing social and communication skills and so on (Kreijns, 2003). Nevertheless, the quality of learning effects is not always assured due to the negative aspects of small group interactions, such as social pressure, inter- and intragroup aggression or conflict and polarization (Strijbos, 2011). In the research field of Computer Supported Collaborative Learning (CSCL), a number of CSCL systems have been proposed for supporting the learning processes using information communication technologies (Jermann, et al., 2001).

On the other hand, many studies on analyzing the small-group face-to-face interactions have been conducted in the research field of multiparty multimodal interaction (Gatica-Perez, 2011). In these studies, interaction management, internal states, social relationships, and so on have been analyzed and modeled based on integrating information via multimodal verbal and non-verbal communication channels such as utterance, gaze and gesture. These findings demonstrate the potential to develop novel CSCL systems that can analyze, assess and also intervene in various learning situations in real time.

However, there is no practical CSCL system embedding these findings in the field of multiparty multimodal interaction. One of the reasons underlying such situation is the lack of applicable platform for developing CSCL systems that can deal with various verbal and non-verbal information (multimodal information).

In this study, we aim to propose a verbal and non-verbal aware platform for developing CSCL systems. The proposed platform is intended to equip a fundamental infrastructure required for any CSCL systems, e.g., session management, and allow developers to implement/extend learning support
tools that can handle nonverbal as well as verbal information provided by the platform to facilitate fruitful communication during collaborative learning processes.

This paper is structured as follows: in section 2, we specify the requirements for developing the verbal and non-verbal aware platform; in section 3, we explain the architecture of the platform with its design principle; in section 4, we discuss the usefulness of the platform by showing an example of learning support tool developed on the platform; in section 5, we introduce some related works and argue the potential of the platform as a verbal and nonverbal aware CSCL system development environment.

2. Requirements

Figure 1 overviews our platform for CSCL systems development. This platform is equipped with various sensing devices in order to capture several verbal and non-verbal information of participants in a collaborative learning (Fig.1(a)). It also provides a fundamental infrastructure for developing CSCL systems, i.e., network and session management (Fig.1(b)). Furthermore, it provides a framework, which allows developers to specify rules to interpret sensed verbal and non-verbal ‘primitive’ information into ‘context’ information among collaborative learners (Fig.1(c)). Consequently, CSCL system developers can concentrate on developing their learning support tools with multimodal interpretation processing (Fig.1(d)) as well as specifying interpretation rules without getting involved in time consuming work for implementing lower level processing.

In order to realize the platform, the following two major requirements must be satisfied:

\[ R1: \text{A mechanism to provide several kinds of primitive verbal and non-verbal information which is the basis for multimodal interpretation (context information).} \]

\[ R2: \text{A mechanism for developers to define learning support tool specific information types (message types), and properly make them communicate in parallel.} \]

2.1 Requirement for Multimodal Interpretation Processing

In order to facilitate the analysis and understanding of conversational structures in multiparty multimodal interaction, Sumi et al. (2011) proposed a layered analysis model as shown in Table 1. The model represents four types of layers combining simple verbal and non-verbal communication signals exchanged among participants in order to achieve multimodal interpretation processing that elicits contextual information such as dominant level transition or participants’ motivation towards their

<table>
<thead>
<tr>
<th>Layer</th>
<th>Summary</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Context</td>
<td>the flow of interaction</td>
<td>dominant level transition</td>
</tr>
<tr>
<td>Interaction Event</td>
<td>the combinations of multiple primitive data</td>
<td>joint attention</td>
</tr>
<tr>
<td>Interaction Primitive</td>
<td>a single motion by a human</td>
<td>looking, speaking</td>
</tr>
</tbody>
</table>
interactions. According to the model, it is first necessary to capture raw data series such as voice, head-movement and eye-coordinates data (raw data layer). From the raw data series, interaction primitive elements such as who is speaking, who is gazing at whom and who is writing (interaction primitive layer) are extracted. The combination of interaction primitive elements leads to identifying interaction events such as joint attention or mutual gaze (interaction event layer). Based on the interaction events, one may infer higher-level contextual interpretation (interaction context layer).

In this research, we take the model as a grain size of layered concept for multimodal interpretation. In order for CSCL developers to build a system which equips such multimodal interpretation mechanism, it is required to pursue a stepwise processing beginning with designing learning activities, preparing an environment for detecting raw data from several sensing devices, extracting interaction primitives from the data, and interpreting them as interaction events.

In this research, in order to reduce CSCL developers’ huge amount of workload, we propose a platform that allows them to access multimodal information without having to care of implementing the detection processing. Here, it is notable that the interaction elements focused on by developers depend on the nature of the collaborative learning and the learning subjects, and the way to deal with the detected data varies according to learning support tools. With keeping in mind this, our platform provides primitive information corresponding to raw data and interaction primitive layers in Table 1, which can be considered as the basis for multimodal interpretation (R1).

2.2 Requirement for Developing Learning Support Tools

There are many types of learning support tools used in collaborative learning, e.g., video-chat tool and text-chat tool as a means of communication, web browser tool for gathering information and shared-board tool for graphical representations, slideshow tool for presentations, etc. In addition to these tools, developers might have to develop their own specific learning support tools according to the target learning activity and subject.

In order to make developed learning support tools run in a network environment on the platform, it is necessary to equip the platform with a communication mechanism which handles various types of messages including learning support tools’ specific ones during the learning activity. For example, the platform needs a specification about how to handle sending/receiving of each message from the learning support tool, meaning that input text messages in the case of text-chat tool, and drawing coordinates messages in the case of shared-board tool, need to be properly specified. Hence, the platform should have a mechanism to communicate not only pre-specified information for authentication and raw level multimodal information captured by sensing devices, but also learning support tools’ specific messages defined by developers (R2).

3. Platform for Developing CSCL Systems

3.1 Platform Architecture

Figure 2 illustrates the architecture of our platform. We employ a client-server architecture style to connect learning support tools used in a collaborative learning session. In the platform, message communication modules in client and server side function in synergy to ensure an adequate distribution of data to the requesting learning support tools.

On the server side, a relational database (CSCL Database) is equipped to store and manage the information about users, sessions, learning history, etc. In addition, to deal with audio and video streaming, the server is implemented by extending Red5 media server (Red5), which supports the real-time messaging protocol (RTMP). Stream communication module distributes audio and video data to the requesting learning support tools. Session management module manages participants’ status and their active learning sessions. CL-data management module registers users’ verbal and non-verbal information and their learning logs sent from clients to CSCL Database.

On the client side, user management module performs authentication processes by communicating with session management module on the server side. Multimodal information management module processes the data stream of the equipped sensing devices in a timely manner, and sends the data to the server. Learning support tool management module manages a group of learning
support tools used in active learning sessions, and distributes messages received from server to the tools properly.

Hereafter, we explain the core functions equipped on the platform in order to satisfy $R1$ & $R2$; in section 3.2, we describe the mechanism that provides verbal and non-verbal information for system developers, and in section 3.3, we provide details about the mechanism that properly distributes messages data exchanged among the server and clients.

### 3.2 Verbal and Non-verbal Information Accessed by Developers

In order to satisfy $R1$, our platform provides developers with several types of verbal and non-verbal information as a basis for multimodal interpretation.

Currently, the platform provides four types of participants’ behavioral information: *utterance*, *gaze*, *writing action* and *head movement*, each of which is often used as a feature to analyze a multiparty multimodal interaction, according to a survey article (Gatica-Perez, 2011) which reviews several topics in multimodal interaction research such as recognizing conversation structure based on speaker-addressee information, estimating roles in a conversation, and identifying change of dominance of a conversation group.

Table 2 summarizes verbal and non-verbal information provided by the platform. Each of the content in the column “Layer” represents the corresponding layer of multimodal interpretation model shown in Table 1. The grain size of the corresponding information is the individual behavior (*raw data / interaction primitive* in Table 1) of a certain participant. Developers are able to access the necessary information in order to interpret them into higher level of multimodal interaction.

1. **Utterance**: *Speech interval* and the *content of utterance* are detected as utterance information. To capture them, a participant’s utterance is recorded via a microphone device. The start and end of an utterance are detected when the audio level exceeds and falls below a certain threshold, respectively. Furthermore, as verbal information, the content of each utterance is provided using a speech-recognition API.

2. **Gaze**: It is well known that the gaze interaction in communication or collaborative activity has several social functions such as expressing one’s intention/feelings and regulating turn-taking (Kendon, 1967). In order to capture the participants’ gaze information, our platform is developed on the premise of using screen-based eye-tracking devices. The platform provides a function for developers to set area-
of-interest (AOI) regions corresponding to each GUI unit such as learning support tools’ window, video object in video-chat tool and text label in text-chat tool. Based on the registered AOI regions, it judges whether the eye movements fall within such AOIs (target objects) at each frame. Developers can access two kinds of gaze information as at the interaction primitive layer: participants’ target objects and respective gazing intervals by just setting AOIs. In addition, our platform also provides eye-coordinates data on the display monitor as at the raw data layer in order for developers to interpret them by defining their own specific rules.

(3) Writing action: Writing action is observed in various context of learning situations such as problem-solving processes, copying down others’ insightful comments, and writing up ideas advanced by participants. In order to incorporate any such writing actions by participants, our platform is designed for handling a digital pen device. The timing of writing information is captured and provided for developers when a participant starts touching and holds off the pen point.

(4) Head movement: In addition to the gaze interaction, the instantaneous reaction by the head movement, e.g. inclining one’s head and nodding, plays an important role to regulate and further the conversation (Kita and Ide, 2007). In order for developers to access such social signal, our platform provides head move information as head direction data (roll, pitch, yaw) in the head-centered coordinate system by using depth camera.

All the verbal and non-verbal information as explained above are detected on each client side in parallel and sent to the server. Developers can access this information by registering target information required for developing their learning support tools.

3.3 Message Processing

In the platform, various types of messages are exchanged among the server and clients via message communication module, e.g., user information for authentication processing, verbal and non-verbal information detected from devices, etc. In order to satisfy R2, we employ a message processing mechanism that discriminates all the messages according to their type. Figure 3 represents the class hierarchy of message classes. All the types of messages inherit ‘msg_type’ property from the ‘Message class’ for identifying their type. In our platform, messages are classified into two categories. One is ‘SystemMessage’ (SY_M) used for authentication processing such as login to or logout to manage a collaborative learning session. We predefined SY_M statically. The other is ‘SessionMessage’ (SE_M) which developers can access in order to develop their specific learning support tools. Moreover, SE_M fall under the following three types:

- **SessionInfoMessage (SI_M):** are generated by the platform when a user participates in or leaves a collaborative learning session. Based on this type of messages, learning support tools can handle who participates in the session.
- **MultimodalDataMessage (MD_M):** correspond to verbal and non-verbal information provided by the platform as shown in Table 2. The data detected from each sensing device is provided for developers as subclass messages of MD_M such as ‘StartWriting’, ‘GazeIn’, etc. These messages are once sent to the server, and distributed to client sides in the collaborative learning session. Developers can register message types of MD_M as necessary for each learning support tool in order to receive data and implement specific multimodal interpretation processing.

320
LearningSupportToolMessage (LT_M): defined by developers when they implement specific learning support tools. Developers define them as subclass messages of LT_M such as ‘ChatTextMessage’, ‘WritingCoordinateMessage’, etc. As same as MD_M, developers can specify their own specific message types of LT_M as necessary for each learning support tool in order to enable specific multimodal interpretation.

To communicate all the messages between the server and clients, they are serialized into JSON format inside the message communication module as shown in Fig.2. Received messages are properly discriminated according to each ‘msg_type’, including those defined by developers at the client side.

4. Usefulness of the Platform

In this section, we discuss the usefulness of our platform. In section 4.1, in order to demonstrate that our platform meets the requirements R1 and R2 and its usefulness, we give a few examples about how learning support tools utilizing verbal and/or non-verbal information provided by the platform are realized. In section 4.2, we illustrate a use case of multimodal aware learning support tools to demonstrate the usefulness of the platform as an execution environment for visualizing data captured in a CSCL session.

4.1 Usefulness from the Viewpoint of Developing Learning Support Tools

Figures 4, 5 and 6 show examples of learning support tools utilizing verbal and/or non-verbal information, implemented on the platform.

- **Talk record tool** (Fig.4): shows a sequence of speech-recognition results corresponding to a participant’s utterance as described in section 3.2. We confirmed that it can be realized by just registering ‘ContentOfUtterance’ (content of utterance) type message provided by the platform as subclass message of MD_M, and implementing the function that appends its values (a couple of a participant’s name and a content) to the result area when the tool deployed at each client receives the message.

- **Non-verbal aware video-chat tool** (Fig.5): the difference between non-verbal aware video-chat tool and ordinary one is that with the former (Non-verbal aware tool), both participants and non-verbal aware tools are aware of other participants’ behaviors such as who is speaking, who is writing as well as who is looking at whom in real-time, whereas with the latter (ordinary tool), only participants are aware of one another without knowing who is looking at whom explicitly. In the platform, audio and video streaming data from microphones and web cameras are sent via the stream communication module. Thus, developers can realize an ordinary video-chat tool easily by just registering SI_M message to receive, detect, login or logout from the session, and implementing the function of displaying/hiding the participant’s video object according to the SI_M. The non-verbal aware video-chat tool can be constructed by extending the basic video-chat tool. It requires for developers first to implement functions for setting an AOI region on each video object, and to register ‘StartGazing’ and ‘StopGazing’ (gazing interval and target object), ‘StartSpeaking’ and ‘StopSpeaking’ (speech interval) as well as ‘StartWriting’ and ‘StopWriting’ (timing of writing) type messages to receive provided non-verbal information as subclass messages of MD_M. Then, they can realize the tool by implementing the following
functions: highlighting the video frame of speaker and displaying labels when others are writing or looking to him/her, in accordance with each \textit{msg\_type}.

- \textit{Gaze aware text-chat tool} (Fig.6): The tool is realized by extending basic text-chat tool so as to be aware of participants’ gazing behaviors. The tool allows participants not only to exchange text messages but also to grasp text messages which participants are focused on at the moment. The basic text-chat tool can be easily built by just defining specific ‘ChatTextMessage’ type message as a subclass of LT\_M which includes each content of the text message, and implementing functions to send and receive the messages of this message type. By extending the basic text-chat tool, we confirmed that the gaze aware text-chat tool can be realized by just doing taking the following routine: implement a function that sets an AOI region on each text-chat message object which is generated when the tool deployed at each client receives this type of message from the server; register ‘StartGazing’ and ‘StopGazing’ (\textit{gazing interval} and \textit{target object}) types message as subclasses of MD\_M. Finally, We could realize the tool by implementing the function that highlights the background of text messages gazed at by other participants. In the case that a text message is gazed at by plural participants, its background is deeply colored to highlight according to the number of gazing participants.

By illustrating the development of learning support tools, we confirmed that the platform satisfies the two requirements: \textit{R2} is satisfied, since SI\_M provided in the platform and LT\_M defined by developers properly communicate through the learning support tools. In addition, \textit{R1} is satisfied, since the developed tools correctly work by receiving (accessing) several kinds of primitive verbal and non-verbal information provided by the platform as subclass messages of MD\_M. Furthermore, we understand the usefulness of the platform for system developers, since we showed that they can realize verbal and non-verbal aware CSCL tools without having to spend time and energy in implementing low level functions for networking and sensor signal processing.

4.2 \textit{Usefulness of Multimodal Aware CSCL Platform}

As a use case of multimodal aware CSCL platform in a practical collaborative learning, we implemented a shared poster tool on the platform. Figure 7 shows the situation where three participants (A who made the poster, B and C) take part in a collaborative learning for discussing about their collaborative research related poster contents using the developed tool. A, B and C can communicate with one another through video-chat (Fig.7(a)), get the control of handling the shared pointer (Fig.7(b)), and point to arbitrary locations to focus on by dragging the mouse pointer (Fig.7(c)). The poster tool also has functions that capture and record participants’ gaze target objects (AOI regions on the poster) and speech intervals of their respective utterances.

Figure 8 shows set AOI regions (colored 17 regions on the poster), participants B and C as viewed by A (Fig.8(a)), speech intervals of each participant’s utterances and gaze information along the timelines captured by the shared poster tool running on the platform. In the timelines (upper right of Fig.8) utterance information is represented as a series of red-colored intervals, each of which corresponds to the participant’s speech interval, whereas upper gazing information of each participant is represented as multi-colored intervals, each of which corresponds to the participant’s gazing interval to an AOI region on the shared poster (the color of the interval is the same as the AOI the learner gaze
at) and finally gray-colored intervals represents the interval during which the participant gaze at other participants’ video object.

While this session took about 30 minutes, participants could focus on their discussion without any disturbance caused by the communication control of the platform. In addition, as shown in the timelines, the platform could properly capture participants’ utterance and gaze information throughout the session. Each of the timelines as in the lower right of Fig.8 represents the situations where a pair of participants (A&B, B&C and C&A) gazed at same sharing objects\(^1\) (AOIs) in the poster simultaneously, which are captured by integrating the two participants’ gaze information. As this shows, they were not always gazing at the same objects, especially at the last half of the session. According to the timelines, we could infer that the participant A first explained her poster contents in a step-by-step manner, while both B and C followed her explanation with gazing at AOI regions on the poster contents (Fig.8(b)). In the end of the session, participants mainly discussed not the poster contents but the future work of the research. This process appears in the timelines (Fig8.(c)); while there is no interval where plural participants were gazing at the same sharing objects on the poster, A was mainly observing the discussion between B and C who actively exchanged their opinions.

As described above, we confirmed that the developed learning support tool successfully run on the platform. Furthermore, we showed the usefulness and potential of the multimodal aware platform for inferring learning processes by utilizing the captured verbal and non-verbal information.

5. Discussion

5.1 Contribution as CSCL Systems Development Platform

As described in the previous sections, our platform make it easier for developers to implement learning support tools which utilize multimodal information through a simple authoring flow consisting in registering target message types which are subclasses of SI_M, MD_M, or LT_M, and adding functions that deal with received messages.

The platform also allows developers to define their unique multimodal interpretation processing by using registered messages, such as to detect situations where a participant is taking notes while another talks by combining utterance and writing, and the situation that who are gazing at same object by referring to plural participants’ gaze targets information as shown in section 4.2.

In total, the platform as a CSCL system development environment contributes to eliminating developers’ workloads to develop their learning support tools, since they do not need any more to care about implementing lower level functions such as multimodal data detection, authentication processing and session management.

5.2 Contribution to the Research Field of Multiparty Multimodal Interaction

In the research field of multiparty multimodal interaction, several studies have been conducted to analyze small-group ‘face-to-face’ interactions based on multimodal verbal and non-verbal information such as speech, gaze and gesture (Gatica-Perez, 2009). Otsuka et al. (2008) proposed an automatic

---

\(^1\) This corresponds to the concept of ‘joint attention’ appearing in face-to-face communication.
identification system that estimates the visual and focus of attention (i.e., “who is looking at whom”), in addition to speaker diarization (i.e., “who is speaking and when”) by using audio and visual signals in real-time. Moreover, McCowan et al. (2004) proposed an automatic analysis method of meeting actions (e.g., monologue, discussions, and presentations) based on an interaction model between participants. In order to achieve this, they extracted a number of simple audio (pitch, energy and speaking rate) and visual (head and hand blobs) data as non-verbal features automatically derived from multiple cameras and microphones, then they employed Hidden Markov Model variations to estimate meeting actions. Hillard et al. (2003) proposed a recognition model of a specific kind of interaction in meetings (agreement vs. disagreement) based on the number of utterances of each participant and positive/negative words included in respective their utterances using machine learning techniques. These studies make practical use of multimodal interaction corpora (Carletta et al., 2005 and Sumi et al., 2011), which are collections of annotated verbal and non-verbal data of multiparty interaction, scientifically analyze and model multiparty human interaction in addition to traditional methods such as participant or video observation. In order to build an interaction corpus, it firstly requires to set the environment where several measurement devices are implemented to collect the intended multimodal data, then annotating each item of data with its proper label.

Our platform makes two major contributions to the research field of multiparty multimodal interaction; The first is that it can capture various verbal and non-verbal information in collaborative learning processes in real-time by setting relatively small-scale equipment (a computer and sensing devices for each participant as shown in Fig.7), and the second is that it makes it possible to utilize captured data including several kinds of primitive verbal and non-verbal information as multimodal interaction corpora for analyzing interaction processes in a collaborative learning situation.

Of course, we need to carefully address whether we can directly apply the findings of multiparty multimodal interaction research to implement functions for analyzing interactions via CSCL systems on our platform, since there are some differences related to the interaction environment (face-to-face vs. remote), the objective of conversations and learning, and the set of tools used during the interactions. Some of the interesting and important future works would be to clarify commonalities and differences between existing findings in multiparty multimodal interaction and the ones we will get through the use of our platform.

5.3 Potential from the Viewpoint of Learning Analytics

Learning analytics (LA) and/or educational data mining (EDM) has recently been the subject of a great deal of attention. The field aims to find out patterns from the big data being accumulated in LMS, CMS and e-learning systems in order to characterize learners’ behaviors and achievements, and make use of them to predict and improve educational functionalities (Peña-Ayala, 2014). El-Halees (2009), for instance, applied data mining techniques called association, classification, clustering and outlier detection rules to the collected students’ data to analyze students’ behavior. Mazza and Milani (2005) proposed a system that visualizes tracking data of students’ behaviors on learning materials, e.g., the history of pages visited, the number of messages read and posted in discussions, to help instructors become aware of what is happening in the learning classes.

In addition to the information used so far in the LA/EDM research field such as logs of learners’ learning contents and scores, our platform can also capture and utilize exhaustive verbal and non-verbal information during the collaborative learning processes as learners’ primitive interaction data. Our platform has a potential to be able to contribute as a learning analytics platform focusing on communication signals level in the interaction processes of collaborative learning, which cannot be captured in traditional LMS and CMS.

6. Conclusion

In this paper, we proposed a verbal and non-verbal aware platform for developing CSCL systems. Our platform is designed to satisfy the following two core requirements: enabling a mechanism that provides verbal and non-verbal information for system developers, and a mechanism that properly communicates various types of data including unique ones defined by developers. It has functions that manage lower level processing such as authentication processing, session management and detection processing of
verbal and non-verbal information, so that developers can concentrate on developing their learning support tools according to the targeted collaborative learning activities. We provided several examples of learning support tools developed on the platform utilizing verbal and/or non-verbal information. Through using them in practical collaborative learning, we confirmed the usefulness of our platform as a CSCL system development environment, and also demonstrated its potential as a basis for learning analytics in computer supported multimodal interaction.

For future works, in parallel to the extension of our platform so as to handle other kind of nonverbal information (e.g., paralanguage information), we intend to perform collaborative learning using CSCL systems implemented on our platform in order to get insights and build up findings about computer supported multimodal interpretations. We also plan to validate the effectiveness of developed learning support tools through operating the platform. It is also an interesting question to make clear how to build learning support tools embedding multimodal interpretation mechanism to capture the interaction context and to support the collaborative learning.

References


Red5, Red5 Media Server, http://red5.org


Using Network-Text Analysis to Characterise Learner Engagement in Active Video Watching

Tobias HECKINGa*, Vania DIMITROVAb, Antonija MITROVICc & H. Ulrich HOPPEa

aCOLLIDE Research Group, University of Duisburg-Essen, Germany
bSchool of Computing and Leeds Institute of Medical Education, University of Leeds, UK
cUniversity of Canterbury, Christchurch, New Zealand
*hecking@collide.info

Abstract: Video is becoming more and more popular as a learning medium in a variety of educational settings, ranging from flipped classrooms to MOOCs to informal learning. The prevailing educational usage of videos is based on watching prepared videos, which calls for accompanying video usage with activities to promote constructive learning. In the Active Video Watching (AVW) approach, learner engagement during video watching is induced via interactive notetaking, similar to video commenting in social video-sharing platforms. This coincides with the JuxtaLearn practice, in which student-created videos were shared on a social networking platform and commented by other students. Drawing on the experience of both AVW and JuxtaLearn, we combine and refine analysis techniques to characterise learner engagement. The approach draws on network-text analysis of learner-generated comments as a basis. This allows for capturing pedagogically relevant aspects of divergence, convergence and (dis-) continuity in textual commenting behaviour related to different learner types. The lexical-semantic analytics approach using learner-generated artefacts provides deep insights into learner engagement. This has broader application in video-based learning environments.

Keywords: video-based learning, learning analytics, network-text analysis

1. Introduction: Video-based Learning and Analytics

Learning by watching videos (Yousef, Chatti and Schroeder, 2014; Vieira, Lopes and Soares, 2014) is becoming more and more popular, especially in new learning contexts, such as flipped classrooms (Kurtz, Tsimerman and Stainer-Lavi, 2014), MOOCs (Guo, Kim and Rubin, 2014; Koedinger et al., 2015), or informal learning. Video watching per se is a passive activity, and therefore it is desirable to provide additional support for learner engagement for better educational benefit (Koedinger et al., 2015, Yousef, Chatti and Schroeder, 2014; Vieira, Lopes and Soares, 2014; Pardo et al., 2015). There are strong indications that increased engagement is more effective for learning. The ICAP framework of Chi and Wylie (2014) can serve as a theoretical reference for this hypothesis. Engagement during video watching can be supported by embedding interactive activities such as quizzes into videos (Giannakos, Sampson and Kidziński, 2016; Klefiodimos and Evangelidis, 2016; Kovacs, 2016; Wachtler et al., 2016), which requires additional effort from the teacher, or by video-annotation features (cf. Chatti et al., 2006) providing opportunities for students to annotate and discuss videos.

Inspired by these approaches, the Active Video Watching (AVW) system (Mitrovic et al., 2016) supports engagement during video watching to facilitate informal learning. The support includes providing micro-scaffolds to facilitate the commenting on videos and the reviewing of comments made by others. Our approach is primarily aimed at informal learning of soft (or “transferable”) skills, such as communicating, negotiating, collaborating, etc. Videos have been shown to be useful for teaching soft skills (Cronin and Cronin, 1992; Conkey et al., 2013), requiring that the learner reflects on his/her own experience and is able to see different perspectives. Another perspective on learning with videos has been adopted by the European project JuxtaLearn (Hoppe et al., 2016): here learners create “dramatised videos” that combine the explanation of science concepts with active storytelling. These learner-generated videos are then shared and discussed on a social media platform.

Drawing on the experience of both AVW and JuxtaLearn, we further explore and refine analytic techniques through which we can characterise the learning benefits, particularly learner engagement around videos. Previous studies performed with the AVW platform focusing on
presentation skills show that only constructive learning results in increased conceptual understanding of the chosen soft skill (Mitrovic et al., 2017). Further analysis of constructive learning behaviour revealed that not all constructive learners increased their domain knowledge, and hence user-adaptive engagement support is needed. Consequently, we characterised constructive learners in three clusters, which informed adaptive support in the form of personalised “nudges” (Dimitrova et al. 2017).

Whereas the studies in the AVW context used data mining techniques related to activity parameters, learner-generated artefacts in the form of textual comments have not been considered. In contrast, the analyses in JuxtaLearn used as a main data source the comments as learner-generated textual artefacts (Daems et al., 2014). This approach was based on network-text analysis (NTA) as introduced by Carley (1997). Here core concepts (terms) are extracted from a given text (a video comment in this case) and arranged in a network. Connections between concepts are introduced based on co-occurrence of the corresponding terms in a window that slides over a normalised version of the text. The central claim of NTA is that the extracted networks are representations of the mental models underlying the texts (Carley, 1997). The transformation of textual artefacts into networks allows for applying network measures and network analysis techniques (Wasserman and Faust, 1994) for further analysis and interpretation of the given data.

Although the cluster analysis of AVW data revealed useful insights for the identification of learner types and the design of adaptive support (see Section 2), it did not reveal much detail regarding the actual differences of the learner types in terms of their commenting behaviour on the content level. Therefore, using only structured interaction log data did not provide sufficient insight into learner engagement with the videos. For such insights, we utilise the learner-generated artefacts. In this paper, we applied NTA to gain deeper insights into the lexical and semantic features of the learner comments. This allowed us to capture aspects of divergence/convergence of vocabulary related to the different learner types. Including temporal aspects, we can also identify continuity vs. variation in textual utterances. This allows for a deeper interpretation of the profiles associated with the learner types, which is a relevant premise to identifying pedagogical challenges and remedial actions.

The rest of the paper is organised as follows. Section 2 elaborates on the AVW system and findings from earlier analyses, including a categorisation of learner types. Section 4 outlines the analysis approach and presents the finding of the network-text analysis of learner comments. Section 5 discusses the findings and points at implications for video-based learning environments.

2. Learning through Active Video Watching: the AVW System

The AVW system is a controlled video-watching environment that supports engagement during video watching via interactive notetaking, tapping into learners’ familiarity with commenting on videos in social networking sites. AVW is customised by the teacher, who selects videos for students to watch, and defines mini-scaffolds for reflective learning. The AVW is particularly aimed at informal learning of soft skills; two studies have been conducted so far focusing on presentation skills and involving university students from engineering subjects (Mitrovic et al., 2016; Mitrovic et al., 2017, Dimitrova et al., 2017). Presentation skills, and transferable skills in general, are highly sought by employers and are crucial for employability (National Research Council, 2012; Walsh and Kotzée, 2010). Teaching soft skills to tertiary students in technical disciplines is challenging, as they are time-consuming and difficult to document (Anthony and Garner, 2016). Learners need to practice under various conditions, receive feedback, reflect on it and do more practice. Teachers typically do not have enough resources to provide such support to each individual student. AVW was developed to address these challenges by providing a video watching space for reflective learning.

Learning in AVW consists of two phases. In Phase 1, students watch and comment on videos individually, using aspects to tag their comments made anytime during the viewing. We selected eight videos from YouTube: four tutorials on how to give presentations, and four example presentations (two TED talks and two 3-minute PhD pitch presentations). The student can stop a video at any time, enter a comment and specify an aspect, which indicates the intention of the comment. For the tutorial videos, aspects aimed at stimulating reflection included: “I didn’t realise I wasn’t doing it” (TA2), “I am rather good at this” (TA3), “I did/saw this in the past” (TA4). There was one additional aspect, “I like this point” (TA1), to encourage the learner to externalize relevant learning points. For the example videos,
the aspects corresponded to presentation skills covered in the tutorials, which included “Delivery” (EA1), “Speech” (EA2), “Structure” (EA3), and “Visual aids” (EA4).

In Phase 2, students review and rate each other’s’ anonymised comments, and can click on ‘view video snippet’ to watch the part of the video to which the comment refers. In such a way, the student can compare his/her own comments to those of others, and further reflect on their experience. The AVW instantiation for presentation skills included five categories for rating comments: “This is useful for me”, “I hadn’t thought of this”, “I didn’t notice this”, “I don’t agree with this”, and “I like this point.”

Figure 1 presents a screenshot from the AVW instantiation for presentation skills, which was used in two studies. The overarching goal of the studies was to investigate whether AVW is beneficial for teaching soft skills. Both studies used the same set of videos, aspects and rating categories, as well as three surveys: (i) prior using the system, participant profiles were collected including demographic information, background experiences, motivation and attitudes using the Motivated Strategies for Learning Questionnaire (MLSQ) (Pintrich and de Groot, 1990) as well as domain knowledge about presentations; (ii) at the end of Phase 1, participants’ knowledge of presentations was checked again, together with questionnaires measuring the users’ cognitive load and perceived usefulness of the system (Hart, 2006; Davis, 1989); (iii) at the end, knowledge of presentations was tested again, and the system’s cognitive load and perceived usefulness assessed again.

One of the AVW studies was conducted with undergraduate students at the University of Canterbury. It compared an experimental condition with reflection mini-scaffolds to the control condition that followed free YouTube-like video watching (Mitrovic et al, 2017). We found a significant increase in conceptual knowledge in the experimental group participants using mini-scaffolds when constructive learning behaviour was followed (i.e. active video watching by making comments and rating others’ comments). There was no significant increase in conceptual knowledge in the control group. The other study, performed with postgraduate students from the Universities of Leeds and Canterbury, looked in depth into constructive learning behaviour (Dimitrova et al, 2017). Thirty-eight out of 48 participants completed all surveys and commented on videos. The initial analyses showed relatively high level of engagement: the participants made a total of 744 comments, and 2,706 ratings (Mitrovic et al., 2016). There were no significant differences between participants based on their gender, age or whether or not they were native English speakers.
Although all 38 participants were constructive learners, not all of them increased their conceptual knowledge after interacting with AVW. Clustering the participants based on their profiles, using k-means clustering, revealed three distinctive types of behaviours generated using the following variables: experience with giving presentations, using YouTube for learning, six MSLQ variables (self-efficacy, extrinsic motivation, academic control, rehearsal, self-regulation, organization), and conceptual knowledge score (Dimitrova et al., 2017). The significant differences between the learner clusters were identified using the 2-sided Kruskal-Wallis test (pairwise comparison with a Bonferroni correction). The seventeen learners in the Cluster 1 exhibited Parochial Learning behaviour. They made relatively high numbers of comments/ratings, had the least presentation experience overall, and had generally low self-regulation and learning skills (they had the lowest MSLQ scores for self-efficacy, extrinsic motivation, rehearsal, self-regulation and organization). Surprisingly, they found AVW the most useful, yet there was no significant improvement of their conceptual knowledge. Cluster 2 exhibited Habitual Video Watching behaviour. These learners were confident, self-regulated students who made fewer comments that the other clusters. At the same time, their conceptual knowledge at the start of the study was the lowest, and there was only a slight increase after using the system. We hypothesised that these participants might be used to watching videos in a passive way so they did not engage sufficiently. Cluster 3 exhibited Engaged Self-regulated (SR) Learning behaviour. This was the “ideal” cluster illustrating the target user behaviour with AVW. The participants were actively engaged while watching the videos, making the highest number of comments and receiving the highest number of ratings on their comments. This cluster was the highest on previous experience and conceptual knowledge on the pretest, and lowest on using YouTube for learning. They significantly improved their conceptual knowledge scores after using the system.

Although the clusters enable characterizing constructive learning behaviour, they do not provide sufficient insights to understand what might be the users’ attention while interacting with videos. E.g., do learners notice relevant points in the videos, are there any notable differences in attention between the three clusters, does engagement change with time, are there any notable links between learners? To answer these questions, and inform the design of intelligent scaffolding to facilitate active video watching, we used the learners’ artefacts generated during the interaction (i.e. the textual comments) and employed networked text analysis. The results are presented in the next section.

3. Computational Analysis of AVW Comments

3.1 Approach

As the first step, a taxonomy of domain keywords in comments was derived using a semi-automatic ontology engineering process. A middle-up ontology authoring approach (Uschold and Gruninger, 1996) was followed, starting from the learners’ answers to questions about conceptual knowledge (bottom-up) and using key categories from several university guides on presentation skills (top-down). Three main categories of domain terms, related to the domain knowledge captured in the tutorial videos, were identified: structure, delivery and speech, and visual aids. For each category, the relevant terms were identified manually by three annotators working independently. As a start, a subset of comments was marked, the disagreements were discussed, and a unified approach for term selection was agreed. The domain terms in all answers to conceptual knowledge (38 students multiplied by 3 surveys for each student) were marked independently by each marker. The majority voting was used to select the relevant terms (a term in a participant’s answer was seen as relevant if it was selected by at least two markers). All cases when there was no majority were examined by a fourth marker, who made changes to the term list. The list of terms was then used in text analysis of the user comments generated during the interaction in the AVW system (Dimitrova et al., 2017). The final refinement of the term list was made by adding the most frequent unigrams in the user comments which were missing from the original term list. This process resulted in a taxonomy of domain terms, including three upper level categories (structure, delivery and speech, visual aids), and the list of frequent domain terms.

Taking the taxonomy as a controlled vocabulary, the learner comments can be transformed into a bipartite learner-keyword network, similar to the approach used by Hoppe et al. (2016). Each learner is connected to all vocabulary terms occurring in at least one of his/her comments. Furthermore, the edges of the resulting network are annotated with further information, in particular, the time when
the relation between the learner and the keyword was established, the learner’s cluster (see Section 2), and a list of video types (example video or tutorial) where the learner used the keyword. The edge-annotated learner-keyword network can then easily be sampled into sub-networks. For example, the sub-network corresponding to the keyword affiliations of parochial learners to the first minute of tutorial videos can be derived by deleting all the edges from the original network that are not annotated by the corresponding cluster, video type, and timestamp. After that, nodes that became isolated through the edge deletions are deleted as well.

The network representation of vocabulary usage during video watching has several advantages. First, it is easy to identify central concepts that are used by many learners, or learners with a broad range of vocabulary terms by calculating the degree centrality of keyword or learner nodes respectively. Furthermore, structural properties of the network, such as the emergence of densely connected regions (or network subgroups) of learners and keywords indicate differences in the vocabulary usage, and thus, attention on different video aspects of subsets of learners.

### 3.2 Usage of Domain Vocabulary

Table 1 gives the most frequent keywords for each learner cluster. The values for the keywords correspond to the fraction of cluster members who used the word at least once, in particular the average degree of a keyword in the corresponding user-keyword network. The threshold was set to 0.5 meaning that all terms in the table column corresponding to a learner cluster were used by at least half of the cluster members. Terms like “presentation” and “story” are frequently used across different clusters, which is not surprising given the topic of the videos. In contrast, differences can be seen in the number of terms and their semantic orientation. Habitual video watchers do not have many shared terms. There are 12 terms used by more than the half of all Cluster 3 participants and the top 6 terms were used by at least 65%, which indicates that these learners tend to use a common vocabulary when commenting on videos. This observation will be further explained in Section 3.3.

<table>
<thead>
<tr>
<th>Parochial learners (14)</th>
<th>Habitual video watchers (7)</th>
<th>Engaged SR learners (17)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keyword</strong></td>
<td><strong>Usage</strong></td>
<td><strong>Keyword</strong></td>
</tr>
<tr>
<td>presentation</td>
<td>0.71</td>
<td>presentation</td>
</tr>
<tr>
<td>story</td>
<td>0.71</td>
<td>pen</td>
</tr>
<tr>
<td>end</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>clear</td>
<td>0.64</td>
<td>line</td>
</tr>
<tr>
<td>beginning</td>
<td>0.57</td>
<td>pen</td>
</tr>
<tr>
<td>talk</td>
<td>0.57</td>
<td>art</td>
</tr>
<tr>
<td>speech</td>
<td>0.5</td>
<td>beginning</td>
</tr>
<tr>
<td>pen</td>
<td>0.5</td>
<td>end</td>
</tr>
<tr>
<td>art</td>
<td>0.5</td>
<td>interest</td>
</tr>
<tr>
<td>slide</td>
<td>0.5</td>
<td>interesting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>text</td>
</tr>
<tr>
<td></td>
<td></td>
<td>eye contact</td>
</tr>
</tbody>
</table>

Table 1 gives a general view of the usage of domain vocabulary for all videos. A significant difference with respect to the used domain terms between example and tutorial videos could be found by a g-test \((G = 318.4, df = 216, p << 0.005)\). The g-test is an alternative to the well-known \(\chi^2\) test which has become popular for computational linguistics since it can also be applied to compare two sparse term frequency vectors (Dunning, 1993). This difference can be especially observed within the learner clusters of parochial learners \((G=189.87, df=162, p=0.066)\) and self-regulated learners \((G=225.7, df=177, p=0.007)\).

By taking a deeper look into the comments made on example/tutorial videos, one can see that affirmative comments such as “Good idea” or “Very interesting” are more frequent in example videos, while especially SR learners post more comments regarding the concrete video content in tutorial
videos. Furthermore, in tutorial videos SR learners show the most agreement on the vocabulary, having 6 terms used by more than the half of all learners in this cluster.

3.3 Agreement on Vocabulary and Shared Attention

The existence of a shared vocabulary in the absence of direct interactions, especially found in the cluster of SR learners (described in the previous section), is investigated in more detail in the following. Again, the analyses are based on the bipartite learner-keyword subnetworks described earlier. Bipartite modularity optimisation (Hecking, Steinert, Gohnert and Hoppe, 2014) was applied to identify densely connected regions (modules) in these networks. Learners and terms are assembled to modules such that within one module the learners are densely connected to a set of vocabulary terms, while the number of edges between modules is minimised. Each module, consequently, represents a set of learners who share a set of terms in their comments. Examples of such network partitions are depicted in Figure 2.

![Figure 2](image)

**Figure 2.** Bi-partite sub-network of learners and vocabulary terms with high modularity (left: habitual video watchers) and low modularity (self-regulated learners).

By definition, modularity optimisation methods create a network partition, i.e., they assign each node to exactly one module, even though possibly this separation may not be very strong for the given network structure. The bipartite modularity (Barber, 2007) is a quality function, which measures how separated the modules of a given partitioning are. It takes the values -0.5 at minimum, 0 for a random partitioning, and 1 in case of perfect separation. Consequently, a low modularity for the identified modules in the learner-keyword networks indicates that a high number of learners and keywords cannot be clearly separated into different modules. Positively speaking, this means that there is a certain degree of common ground in terms of shared vocabulary between these actors. This can especially be observed for the SR learner cluster (right-hand side of Figure 2). The corresponding modularity values can be found in Table 2. Here, the 17 learners were split into 7 clusters for all videos and for the tutorial videos respectively. It can be seen that each module has some characteristic terms that are not used by the learners in other modules.

However, there is also a high share of terms that have many connections to learners of different modules, and thus, cannot be clearly assigned to a particular module. This gives further evidence that the SR learners have a certain agreement on the vocabulary used in their comments. The reason can be that the SR learners follow the videos thoughtfully and take up concepts from the videos in their postings. Particularly in video tutorials, it could be observed that these learners tend to post comments on the actual video content. In contrast, the habitual video watchers and parochial learners show a different behaviour. Since these learners tend to post more affirmative comments on the general style of the videos, their corresponding learner-keyword networks can be split into more separated modules, which results into higher modularity values (Table 2). The habitual video watchers denote an extreme
case where each of the seven members of this cluster forms an own module in the networks (Table 2 and left-hand side of Figure 1).

Table 2: Characteristics of partitioned networks extracted from all videos vs. tutorials for different learner clusters

<table>
<thead>
<tr>
<th>Cluster: Video type</th>
<th>No. modules</th>
<th>Modularity</th>
<th>Keywords / user</th>
<th>Users / keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parochial: all videos</td>
<td>5</td>
<td>0.37</td>
<td>28.6</td>
<td>28.57</td>
</tr>
<tr>
<td>Parochial: tutorials</td>
<td>6</td>
<td>0.46</td>
<td>18.1</td>
<td>18.08</td>
</tr>
<tr>
<td>Habitual: all videos</td>
<td>7</td>
<td>0.45</td>
<td>18.43</td>
<td>18.43</td>
</tr>
<tr>
<td>Habitual: tutorials</td>
<td>7</td>
<td>0.46</td>
<td>14.43</td>
<td>14.48</td>
</tr>
<tr>
<td>Self-reg.: all videos</td>
<td>7</td>
<td>0.3</td>
<td>34.7</td>
<td>34.7</td>
</tr>
<tr>
<td>Self-reg.: tutorials</td>
<td>6</td>
<td>0.35</td>
<td>22.41</td>
<td>22.41</td>
</tr>
</tbody>
</table>

3.4 Attention Shift During Video Watching

In this section, the joint attention of the three learner clusters is analysed on a fine-grained level. The domain vocabulary terms used by the students in each minute of the particular video are investigated in order to identify different patterns that further characterise the engagement of different learner clusters. On the one hand, periods of high attention indicated by a high number of vocabulary terms per learner have to be identified. On the other hand, it is of interest how the used vocabulary changes during the course of the video.

Figure 3. Attention diagrams for two example videos for different learner types
We developed an integrated visualisation that captures these aspects (Figure 3). The number of vocabulary terms per learner in every minute of a video is represented by the size of the circles on the horizontal axis. The more vocabulary terms have been used (indicating higher attention), the bigger the diameter of the circle is. The width of the horizontal arcs depicts the overlap of vocabulary terms in two consecutive minutes so that attention shift becomes visible. The terms in the circles are those that were used by at least two learners in the corresponding minute.

Figure 3 shows the attention diagrams for two videos (one tutorial and one example), to illustrate different patterns of attention shift for different learner clusters. The typical pattern for parochial video watchers is the high frequency of general and affirmative comments, as already stated above. These types of comments cannot be clearly attributed to a part of the video since video content is not mentioned. This explains the relatively high overlap in the used terminology. It can also be seen that the use of vocabulary terms is more or less evenly spread over the course of the videos. Habitual video watchers do not only have a low usage of domain terms, but also show discontinuous posting activity in certain minutes of the videos. Here it is important to mention that this cluster comprises of only seven learners, which can also partially explain the discontinuation of the posting activities. The SR learners show activity throughout the video, similar to the parochial learner cluster. However, there is a general tendency to write comments at the beginning of a video. In the tutorials, the usage of vocabulary terms is more oriented towards the content of the video indicating that the commenting activity of self-regulated learners is more guided by the topics discussed by the tutorials, and consequently, there is less continuity (or overlap) in the used terms in consecutive video minutes.

4. Discussion and Conclusion

In this paper, we applied the network-text analysis of video comments to investigate engagement related to attention and thematic focus of learners in active video watching tasks, where learners were supposed to watch examples and tutorials on giving presentations, and additionally post comments on specific parts of the videos. Earlier analysis (Dimitrova et al., 2017) showed that learners were subdivided into three different clusters with respect to their constructive learning behaviour and video watching habits. The overall goal of the lexical semantic analysis conducted in this paper was to gain further insights by having a closer look into actual content produced by these clusters, in particular, the usage of domain vocabulary in video comments. We believe that only such a combination of behaviour and content analysis for revealing characteristic patterns of engagement and attention can support tutors and designers of video-based learning scenarios in designing good videos. Furthermore, this can give indicators which guidance mechanisms could be established to improve the learning experience in active video watching tasks. Network-text analysis is especially suited to achieve these goals since it allows for extracting the overall relational structure of learners.

**Insights gathered from NTA.** Applying NTA, we were able to extract a lexical-semantic structure from the learner-generated artefacts in the form of unstructured textual comments. This provided an interpretable model analysed with well-established network analysis methods. There were several important insights which were not identified by the earlier analysis using only interaction logs.

Firstly, it was shown that the **usage of vocabulary terms differed significantly between example videos and tutorials**. Tutorial videos seemed to be better suited for engaging learners in reasoning about specific concepts, while examples trigger more comments referring to presentations in general, like “interesting”, or “speech”. The highest agreement on a common set of terms, and therefore, the existence of joint attention could be found for self-directed learners, while parochial and habitual video watchers only have very few keywords in common.

Secondly, the observation of the **emergence of a shared vocabulary was confirmed** based on a fine-grained analysis. The usage of domain vocabulary was investigated by identifying densely connected modules of learners and used terms in learner-keyword networks derived from the video comments. In particular, the identified network modules corresponding to the self-directed learners could be characterised by some unique terms, but in addition, there was also a high share of keywords that were used by almost all modules. This finding is interesting for further research on video-based learning, since it indicates that engaged self-directed learners are able to recognise important concepts from the videos and use them in their comments. Thus, it will be easier for these learners to find a
common ground or at least common vocabulary in possible post-video group discussions. For other types of learners, this phenomenon was not very salient, especially for habitual video watchers with little background knowledge. This indicates that additional scaffolding would be needed for these learner types to point at important aspects in the videos.

Thirdly, we observed that the attention of learners shifted during the course of a video. It could be shown that the commenting behaviour and attention for particular learners differed between the three learner clusters. The parochial and self-directed learners tend to post comments throughout the entire duration of the videos. However, while parochial learners mainly write general affirmative comments and opinions, self-directed learners were more guided by the actual video content since their comments were more closely aligned to concepts discussed in specific moments of the video, especially in tutorials. Habitual video watchers show a very different pattern - little activity and posting comments for much narrower periods of the video.

Implications to video-based learning environments. As a possible consequence of the presented findings, future active video watching tasks could be enriched by scaffolding mechanisms. This can help learners to focus on important aspects and to find a vocabulary to express themselves in video comments more precisely. Thereby, the emergence of a shared vocabulary is desirable since this would be beneficial for post-video discussions and facilitate conceptual framing. This kind of lexical-semantic support can be achieved, for example, by presenting a list of important domain vocabulary terms to the learners. This would, on the one hand, guide their attention to important aspects discussed in the video (as identified in the comments), and on the other hand, help learners to find the wording for dedicated comments (by pointing at example comments).

Since the more specific comments could be found in tutorial videos that had to be watched before the example videos, it can be assumed that some of the learners had difficulties to apply the concepts mentioned explicitly in the tutorials when they commented on examples. By using NTA, relevant terms extracted from tutorial comments could be presented to the learner in example videos, which can support learners to put example video content in a conceptual framework. This can lead to more specific comments, helpful for triggering reasoning about the learning objective than simple affirmative comments.

The automatic analysis of learner-generated artefacts to gain an understanding of learner engagement with videos is applicable in a broader context. Starting from textual comments, it will be possible to derive notable links between learners and learner behaviour, to identify the areas of attention for a group of learners or for a specific learner, and to depict how attention changes during the video. We intend to apply NTA on a recently completed large user study with engineering students at the University of Canterbury. Furthermore, we will investigate the application of AVW in other domains where videos are used as part of soft skill training, e.g. communication skills in Medicine. This will allow us to validate/tune the learner clusters and derive lexical semantic characteristics for each cluster.

Acknowledgements

This research was supported by the EU-FP7-ICT-257184 ImREAL grant, a teaching development grant from the University of Canterbury, and an Ako Aotearoa grant. We thank Lydia Lau and Amali Weerasinghe for their contribution to designing and conducting the user study, and we thank all participants in the experimental study. Regarding the analysis methodology, we acknowledge to have drawn on results from the EU project JuxtaLearn (EU-FP7-ICT-317964).

References


Koedinger, K.R., Kim, J., Jia Z., McLaughlin E., Bier, N. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. *Proc. 2nd Learning @ Scale*, pp. 111-120.


Synergizing Online Group Knowledge

Yen-An SHIH & Ben CHANG*
Graduate Institute of Learning and Instruction, National Central University, Taiwan
*bchang.tw@gmail.com

Abstract: A huge amount of users’ data are created on the Internet day by day. How to synergize those users’ data to provide useful information will be an important issue. Based on the phenomenon mentioned above, this study aims to refine the users’ data on an experimental social network platform named CoCoing.info to a reintegrated concept map. The CoCoing.info platform has been online for a half year. During the period, 646 users enrolled, and 2,096 concept maps and 4,569 users’ responses were created. To synergize those users’ data into a connected concept map for further value-added applications, two computer algorithms, term-generation and term-association programs, were designed to filter the users’ data and to reintegrate the filtered data. By the term-generation algorithm, 46,490 terms were filtered. Moreover, the term-association algorithm synergized as an 18,011-node concept map based on the filtered terms. The synergized concept map can potentially provide users to implement adaptive value-added applications on it.

Keywords: Learning analytics, big data, knowledge building, meta-knowledge

1. Introduction

With technology evolution, traditional web-based learning environment, such as MOODLE usage, has been getting challenges since general students use more social network applications to interact with their peers rather than the teacher-led activities. Moreover, with the Web 2.0 (O'Reilly, 2005) design method, those students, they can create a lot of content on their owned social network and give responses instantly. The student-created content has great potential values if we can apply useful technology to analyze and to synergize those users’ data. Such kind of design thinking has been widely discussed in some research (Chatti et al., 2012; Ferguson, 2012), like big data research (Mayer-Schönberger & Cukier, 2013), learning analytics (LA) and educational data mining (EDM).

Nowadays, many popular and prevailing knowledge community (Scardamalia & Bereiter, 2006) were created. Wikipedia is one of the well-known platforms, which provides online services for users to create, edit, organize, and distribute knowledge. The purpose of knowledge community is to provide the open-sourced knowledge access and fair discussion to the Internet users. They may come from different backgrounds and abilities, but have the same study issue and come together to maintain the knowledge. They build the structural knowledge, propose innovative concepts, update the last idea, and post questions. In such kind of knowledge community, the users’ work is primarily valued for what it contributes to the knowledge group (Scardamalia & Bereiter, 2014). Comparing to individual learning, knowledge community focus on learners’ contribution, interaction and collaboration work.

A huge amount of knowledge is created every day on online knowledge communities. Hence, an online open-source knowledge community must handle the enormous quantity of data with the huge storage capacity. The circumstance brings and opens the era of big data in education. That means the learning process, path and pattern also can be recorded when learners are learning. By the recording data, we can conduct learning analytics method and figure out each learner’s personal learning profile, for example, (1) what subjects are student’s expertise, (2) what is student’s interests, (3) what things student likes or dislikes, (4) and how does a student interact with knowledge community. However, before this, how to collect the huge data from learners and analyze the data become an important issue.

To sum above, in the web 2.0 environment, the volume of data is increasing faster than before. On the online knowledge community, the knowledge and learning data can automatically be saved in the database. Through learning analytics, the online knowledge data can be reintegrated and processed, and the result can further be provided to the learners. However, only a few prior studies discussed (1) how an online knowledge community can be designed to collect learner’s knowledge data, (2) how to...
utilize the data, and (3) what is the usage of the analysis output on supporting learner’s learning. To this end, this study aims to examine:

(1) How to design a social network platform in which the students can organize their personal social network, and have their learning activities on the platform as well?
(2) How to retrieve the concepts created by the students, and convert the retrieved concepts into a reintegrated concept map?

2. Related Work

2.1 Collaborative Knowledge Creation

The new communication technology provides the potential web-based environment for the students to learn on the Internet, and the online resources can be widely accessed via many portal websites, and the content can be supported with multimedia such as image, audio, video and flash animation to give learners more information than traditional text content. On the websites, the online learning content is not only created by teachers, Web 2.0 allows students from individual learning to collaborative create knowledge, share ideas and discuss with peers through online tools on a learning community. The handy usage of knowledge distribution and access triggers the trend that student has interest and engages in creating and sharing personal material. A student in the online learning community transfers the role from traditional knowledge receiver to knowledge creator and information provider. Meanwhile, the learning content is also widely accessed and evaluated by others, each learner can provide feedback and enhance the quality of knowledge contribution by discussion (Yücel & Usluel, 2016). The online environment improves the traditional teaching method and provides a space for the distributed team work to collaboratively contribute and maintain knowledge via technology device.

Past studies indicated that online collaborative knowledge work could provide the chance which teacher and students can create knowledge without the limitation of time and space (Shukor, Tasir, Van der Meijden, & Harun, 2014; Lampe, Wohn, Vitak, Ellison, & Wash, 2011). Such collaborative learning environment motivates students to create more knowledge and personal opinion because peers play an important role as guide of knowledge creation (Yücel & Usluel, 2016). In the Liaw, Chen, and Huang’s study (2008), researchers developed a web-based collaborative learning system for students to learn with the online learning material and share their knowledge which relates the learning content and to examine their attitudes. The result indicated that collaborative learning environment may promote the efficiency of knowledge creation and students actively share ideas. To sum up, the collaborative knowledge creation environment has many advantages on improving student’s learning performance. However, few studies discussed how to construct a collaborative learning environment and manage the knowledge data. To fill this gap, this study developed an online platform called “CoCoing.info” to implement the collaborative learning environment and to examine the effectiveness.

2.2 Educational Data Mining

Educational data mining (EDM) has been widely discussed since the mature development of computer science. EDM is a multidisciplinary research field in terms of using the technics of statistics, machine learning and data mining (Dutt, Ismail, & Herawan, 2017). EDM relies on computing power to deal with the processing of data collection, data extraction, analysis, and even system training with the large volume of data. EDM focuses on the analysis of profiles and students’ learning data in which learner acts on the digital learning environment in order to discover students’ learning process and profile (Levy & Wilensky, 2011). The analysis result can be used in monitoring students’ learning status as well as providing feedback to the student for improving the learning problems, and to understand how and what student learns in that environment. EDM not only enriches the value of educational data by the contribution of learners who have different backgrounds but also provides the adaptive assistance to each learner according to s/he learning status.

There are many examples of the applications of EDM. For instance, students’ plagiaristic behavior on assignments is a challenging problem for the teacher because it is cumbersome to examine
by a human being. To solve this problem, Akçapınar (2015) applied the EDM technic in text mining on the volume of 4268 reflection texts with 59 participants. The plagiaristic ratio was provided to students as a reminder while they were posting. There is a significant improvement on the student’s plagiaristic behavior according to the result of the experiment. Romero, Ventura, & García (2008) introduced how to apply the EDM on mining Moodle data. Some studies used EDM on adaptive learning with a proposed data mining system (Lin, Yeh, Hung, & Chang, 2013; Romero, Ventura, Zafra, & De Bra, 2009). However, only few research discussed how the implement the EDM on the integrating the group knowledge and discuss the mining process. To fill this gap, this research aims to process of how to create the knowledge relation by text mining and generating the integrating concept map.

3. Data Source: CoCoing.info Social Network Platform

To exam the goals of how to reintegrate group knowledge, an online learning platform called CoCoing.info had been developed to collect the knowledge data from learners and conducted in the experiment. As shown in Figure 1, students contribute knowledge and learn in the ubiquitous computing environment. The knowledge data was converted into terms by the term-generation algorithm and saved in the database. Those terms were further used to connect concepts and knowledge by the term-association algorithm and create a reintegrated concept map to support further value-added applications. The scenario of reintegrating group knowledge on CoCoing.info is shown in Figure 1.

3.1 CoCoing.info Background Introduction

There are two major designs on CoCoing.info for facilitating learners producing knowledge and cooperative constructing ideas. The first major design is the concept construction, and the other is the social learning network design. The structure of CoCoing.info is shown in Figure 2.
3.1.1 Concept Construction

Knowledge is a thinking product which exists in the personal mind. However, how to transform the invisible thinking into the visible and structured knowledge is a challenge. In this study, we utilized the concept map tool and implemented into CoCoing.info platform for facilitating learners building knowledge. The node and line are two basic elements in a concept map. With the main idea, it can be expanded to many nodes and connected by lines. CoCoing.info provides functions for learner easily building a structured concept map. By the clicking function buttons, the learner can create the many colorful nodes and links on a map, and attach relative images and files as learning resources on each node to illustrate the learning content. Furthermore, a node can connect to online learning resource with URL so that learner can easily integrate online learning material into a concept map. The concept map construction page and the description of each function are shown in Figure 3.
3.1.2 Social Learning Network

The social network is an important element among students when they are learning. The relationship between peers is a pattern of how students interact and influence others. On the CoCoing.info platform, there are several functions for learners to create and represent their social network from the classroom. For example, they can add their peers as a friend by name or email on CoCoing.info, and two or more friends can gather and form a learning group. They can become the friend relationship with each other. The social network among students is a complicated net connection working on CoCoing.info. With the integration of social network and concept map, CoCoing.info can support the collaborative learning environment. The learner can share a concept map with friends or groups. Learners can simultaneously and collaboratively construct the same concept map with peers, and teacher can join students and provide guidance when they are constructing. Each node displays the last editor’s name. In the process of concept map constructing, learners and teachers can discuss, provide feedbacks, and exchange ideas simultaneously with each other by the response function.

3.2 Participants

There were 646 users enrolled in the system. They were ten to twenty-four years old covering in elementary school, junior high school, and college students. All of the teachers had been educated how to use the functions on CoCoing.info by researchers with face-to-face teaching. Therefore, students registered their personal account and went through the operation guidance about CoCoing.info over three-hour class periods for two weeks by teachers to ensure they were skilled. In this study, the teachers were as the role of supporter and provided guidance while learners constructing their concept maps.

4. Data Synergy: Term-generation and Term-association Algorithms

4.1 A Process Model for Learning Analytics

On CoCoing.info learning platform, learners can easily create personal and cooperative concept maps, and the concept map data is automatically stored in CoCoing.info database. To achieve the data reusing, we proposed the model of learning analytics and try to mine out the useful information from the concept map data. The analytics output could provide the feedback to students in order to enhance their learning performance and extend their knowledge and perspective. The process mode of learning analytics is shown in Figure 4.

![Figure 4](image_url)

Figure 4. The process model of learning analytics.
4.2 Term-generation Algorithm

In CoCoing.info database, the concept maps and responses data was selected as the resource of term generation and training. A total of 2,096 concept maps and 18,011 nodes which generated by learners without null or empty value from the period of November 1, 2016 to April 30, 2017 were selected as experiment data, and the 4,569 responses data was selected as well. The content of nodes and responses included English and Chinese text. To generate terms, the content was split into a single word as the basic element. Each word went through others and combined as a new phrase. The new phrase may contain two or more word in the phrase. With the loop process of phrase composition, we counted the sum of repetition and recorded it as the weight for each phrase. According to the weight number, a new term may be created if the number is higher than others. Therefore, the new term can be used in analytic and continue changing its weight number by the new data and the loop process. The term extraction algorithm is shown in Table 1.

Table 1: The term-generation algorithm pseudo code.

<table>
<thead>
<tr>
<th>Line</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Do Loop</td>
<td>Extract text from nodes and responses</td>
</tr>
<tr>
<td>2</td>
<td>If source is Chinese characters</td>
<td>Set position at x word</td>
</tr>
<tr>
<td>3</td>
<td>Split text from position x, y words as a term</td>
<td>Filter alphabet from text</td>
</tr>
<tr>
<td>4</td>
<td>Push term into array cell</td>
<td>Combine alphabets as a term</td>
</tr>
<tr>
<td>5</td>
<td>End of loop</td>
<td>Push term into array cell</td>
</tr>
<tr>
<td>6</td>
<td>Save text, array, array length as a record</td>
<td>Save text, array, array length as a record</td>
</tr>
</tbody>
</table>

According to the term-generation algorithm mentioned above, a total of 46,490 terms were generated based on the 2,096 concept maps and 4,569 users’ responses data stored in the CoCoing.info database. Figure 5 displays part of the terms list. In Figure 5, a user’s sentence is parsed to a list of terms for further analysis.

Figure 5. The screen capture of term-generation page.
4.3 Term-association Algorithm

As mentioned above, the relation between nodes is represented as a line in a concept map. Each node may have its parent node and child node. However, the lines in a concept map cannot interact with different concept maps. For this purpose, the terms work as an index which represents the advance relation between nodes if the terms are similar. We compared the term list among two nodes and counted its similarity with the algorithm. There were 46,490 terms generated and the highest occurrence number was counted by 3,176 times. With the highest number, we can calculate the weight for each term by the occurrence number. However, not all of terms can be used in the analysis, the lower weight terms were removed in the pre-processing and only high weight terms were kept to the comparison. The similarity number ranges from 0 “no similarity” to 100 “high similarity”. The association algorithm is shown in Table 2.

Table 2: The term-association algorithm pseudo code.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Collect terms of text A and text B as array A and array B from database</td>
</tr>
<tr>
<td>2</td>
<td>For i= 1 to the length of array A</td>
</tr>
<tr>
<td>3</td>
<td>For j= 1 to the length of array B</td>
</tr>
<tr>
<td>4</td>
<td>Compare term i and term j</td>
</tr>
<tr>
<td>5</td>
<td>If term i and term j are matched</td>
</tr>
<tr>
<td>6</td>
<td>Match count +1</td>
</tr>
<tr>
<td>7</td>
<td>End of For loop</td>
</tr>
<tr>
<td>8</td>
<td>End of For loop</td>
</tr>
<tr>
<td>9</td>
<td>Calculate the minimum of array A and array B</td>
</tr>
<tr>
<td>10</td>
<td>Similarity = match count / minimum</td>
</tr>
<tr>
<td>11</td>
<td>Visualize terms relation based on similarity result</td>
</tr>
</tbody>
</table>

4.4 Terms Relation Visualization: Integrative Concept Maps Generating

By the term-association algorithm, the similarity of the relation among nodes can be calculated. Therefore, the integrative concept map can be generated based on the similarity calculated results. Currently, the reintegrated concept map is an 18,011 nodes concept map which is a huge map. The reintegrated concept map is expanding day by day automatically by the term-generation algorithm and term-association algorithm once users continually use the system. Figure 6 only displays part of the reintegrated concept map. However, the whole reintegrated concept map can be retrieved via the Internet for further value-added applications.

As we can see in Figure 6, the nodes were displayed as the block with light blue color. The concept map relation was represented as the solid line with black color. The nodes were drawn closer if they were in the relation of concept map which showed as the high-density area, and there were few outliers which were deleted and had no relation among nodes as well. Different similarities were drawn with the dotted line with the red color (51~60), orange color (61~70), green color (71~80), blue color (81~90), and pink color (>90), respectively.
5. Discussions and Conclusions

No doubt that on the Internet, the users generate a big volume of data every day. Novel data analysis approaches, such as educational data mining and learning analytics, can retrieve those data to provide useful information. The research trend of applying EDM and LA in the educational system is getting important (Romero & Ventura, 2007).

In this study, an online platform named CoCoing.info was implemented which is applied to study on how to design an online knowledge community, and how to collect the structured users’ data. Meanwhile, two computer algorithms, term-generation and term-association, were adopted to filter the users’ data and to reintegrate the filtered data into a reintegrated concept map automatically.

A half year period of the announcement of the CoCoing.info platform, a total of 2,096 concept maps, 18,011 nodes, and 4,569 responses were collected in the database. Moreover, there were 46,490 terms generated by the terms-generation algorithm, and the highest occurrence number in terms is 3,176 times. The results indicated that the relative concepts can be organized and connected. Moreover, the concepts from different learners can be combined together for further applications. Those results provide further learning platform designers a guideline and approach on designing adaptive learning, appropriated responses, and users modeling system. More specifically, those adaptive systems mentioned above need a huge concept map as the fundamental knowledge set to provide appropriate reactions, and this study illustrated how such a huge concept map can be created and reintegrated automatically.

Comparing to the big data research, this study still bases on small samples. However, the design has provided a useful reference model and the results have revealed some interesting insights that further designers’ references. Furthermore, the algorithms discussed in this study of meta-knowledge generation can be applied in different tasks in education. For example, the meta-knowledge can provide the learning suggestions for the learner to improve learning performance, and improve teacher’s pedagogy as well. The meta-knowledge also can be used in the training of machine tutor as the scaffolding while student learning.
References


A Presentation Avatar for Self-Review

Keisuke INAZAWA* & Akihiro KASHIHARA
The University of Electro-Communications, Japan
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585
*keisuke.inazawa@uec.ac.jp

Abstract: Presentation is one of the most important activities for researchers. For proper presentation, self-rehearsal is often conducted. In self-rehearsal, presenters usually make presentation to themselves with their PC. However, they often miss finding some points to be modified, since they need to review it concurrently with making their presentation. On the other hand, there is another way for self-reviewing, in which they could make a video of their presentation and then check it out. Although it allows them to direct more efforts to review, they would feel quite uncomfortable due to their looks and voice on the video. Our approach to this issue is to design presentation avatar, which allow the presenters to objectively self-review their own presentation without any uncomfortable sense. In this paper, we propose and demonstrate a presentation avatar, which reproduces the presenters make. This paper also reports a case study using it, in which we compare self-review with presentation video and self-review with presentation avatar. The results suggest that presentation avatar has a potential to promote self-reviewing to improve presentation.

Keywords: Presentation, Avatar, Self-review, Rehearsal, Objective review

1. Introduction

Presentation is one of the most important activities for researchers. They are accordingly required to develop skill in presenting their work properly. Proper presentation needs several rehearsals, which involves reviewing the presented contents, oral explanation, gesture, etc. to improve the presentation. It is crucial particularly for unskilful researchers to conduct presentation rehearsal. Such rehearsal would also contribute to developing presentation skill.

In research laboratory, presentation rehearsal is often conducted in two ways, which are rehearsal with peers (lab members) and self-rehearsal. In rehearsal with peers, presenters could receive peer reviews indicating points to be modified (Ryo Okamoto, and Akihiro Kashihara, 2007). In self-rehearsal, they need to make presentation and review it by themselves from an audience point of view to find the modified points. Self-rehearsal is often conducted before rehearsal with peers.

The main issue addressed in this paper is how to allow unskilful presenters as learners to self-review and improve their presentation. In self-rehearsal, presenters usually make presentation to themselves with their PC. However, they often miss finding some points to be modified, since they need to review it concurrently with making their presentation. On the other hand, there is another way for self-reviewing, in which they could make a video of their presentation and then check it out. Although it allows them to direct more efforts to review, they would feel quite uncomfortable due to their looks and voice on the video. As Holzman et al. pointed out (Holzman, Philip S., and Clyde Rousey, 1966), such uncomfortable sense often occurs from discrepancy between the voice one hears in one’s head and the recorded voice. In the same way, the presenters could feel uncomfortable with discrepancy between the looks one expects and the recorded looks. This uncomfortableness prevents the presenters from objectively self-reviewing their recorded presentation from audience viewpoints.

In order to address this issue, we are designing presentation avatar (P-Avatar for short), which acts as computer agent or educational robot for reproducing the presentation presenters make, and for making presentation as their proxy. The main aim of P-Avatar is to allow the presenters to concentrate on self-review of their own presentation without any uncomfortable sense and to find points to be modified.

In this paper, we propose and demonstrate P-Avatar that reproduces presentation learners make. This paper also reports a case study using it, in which we compare self-review with presentation video.
and self-review with P-Avatar. The results suggest that P-Avatar has a potential to promote self-reviewing to improve presentation.

2. Self-Review

2.1 Presentation Rehearsal

Presentation rehearsal allows learners to gain reviews including points to be modified and to improve their presentation. It also gives them an opportunity to develop their presentation skill. It can be modeled as a cyclic process involving three phases as shown in Figure 1, which are preliminary presentation, review, and modification. In the phase of preliminary presentation, learners rehearse presentation with a presentation document (P-document for short) such as PowerPoint/Keynote one according to the context expected. In the review phase, they check out their presentation with peers including more skillful lab members or by themselves to gain points to be modified. In the phase of modification, the learners are expected to follow the review results to modify the P-document, oral explanation, gesture, etc. By repeating these phases, the learners can improve their presentation before actual presentation.

In developing presentation skill, it is indispensable for unskillful learners to enhance the ability to self-review (Nancy, D., 2008), although peer review is an instructive way for improving presentation. In self-review, learners are required to have an audience view to objectively check how they used P-document, how they made oral explanation and non-verbal behavior such as gesture and facial expression, and also to check the contents of P-document and oral explanation, etc.

Such objective self-review allows them to find out what should be modified. The learners are particularly expected to become aware of excess/deficiency/suitability in the contents, design, and slide order of the P-document. They are also expected to become aware of improper non-verbal behavior including eye contact to the audience, behavioral habits, pointing to the slide, and improper oral explanation including speed, emphasis on keywords/sentences, connection between slides, intonation, intervals, haste, etc.

However, it is not so easy for unskillful learners to gain such awareness of points to be modified from self-review. How to promote self-reviewing to increase awareness of the modified points in presentation is the main issue addressed in this paper.

![Figure 1. Rehearsal model.](image-url)
2.2 Problems in Self-Review

The normal way for learners to do presentation rehearsal involving self-review is to make presentation to themselves with their PC as shown in Figure 2(a). However, it is quite difficult to self-review it in detail since the self-review process is conducted concurrently with presentation. Some points to be modified could be accordingly missed.

On the other hand, there is another way of the self-rehearsal as shown in Figure 2(b), which is to make a video of their presentation and then to check it out. This allows the learners to direct more efforts to review. But, they would feel quite uncomfortable due to their looks and voice on the video. Such uncomfortable sense would occur from discrepancy between the looks/voice one expects in one’s head and the recorded looks/voice. It is quite difficult for the learners to review the presentation objectively with this uncomfortable sense, which prevents them from concentrating on self-review. Even though they could overcome the uncomfortable sense, in addition, there would be a limit to finding points to be modified. Although objective perspective contributes to gaining awareness of points to be modified, it is quite difficult to hold it in self-review.

In order to resolve the above problems, it is necessary to remove the uncomfortable sense to be brought about in self-review to promote checking out the presentation and enhance awareness of the modified points.

2.3 Related Work

Related work on supporting self-rehearsal of presentation has been mainly addressing how to automatically analyze non-verbal behavior and oral explanation from motion capture of presentation. For instance, Kurihara et al. has proposed a system, which analyzes degree of eye contact to audience, oral speed, intonation, and hesitation, from captured motion and then to present the analyzed statistical data to learners as review results during/after their presentation (Kazutaka Kurihara, et al., 2007). The learners could have training of self-rehearsal with the system, and could become aware of non-verbal behavior and oral explanation to be modified with such data. Although such analyzed data seems instructive for presentation improvement, self-review is limited within the non-verbal behavior defined in advance by the system. The system does not provide learners with any support that promotes reviewing their P-documents.

Zhao et al. has also proposed a system, which detects and evaluates non-verbal behavior necessary for proper presentation such as posture, attitude, gesture, etc (Xinbo Zhao, Takaya Yuizono, and Jun Munemori, 2015). If the system detects improper behavior during presentation, learners immediately receive feedback with evaluated information. After presentation, they can also check out the presentation video with the evaluated information. However, they would still have difficulties in gaining awareness of points to be modified since the system provides no support for decreasing uncomfortable sense brought about by looking their looks and listening to their voice in the self-review process.

Chollet et al. has proposed a system using virtual audience (Chollet et al., 2015). The system analyzes eye contact to audience and number of pause fillers with audiovisual sensors, and then presents either directly visualized performance measures or implicit nonverbal feedback from the virtual audience during their presentation. The virtual audience will behave positively when the presenter is performing well, and behave negatively when the presenter is not performing well. For example, they smile when the presenter make eye contact with them, or they look away when a pause filler occurs. In this system, however, it is also difficult to gain awareness of points to be modified in detail without support after presentation.

P-Avatar proposed in this paper does not intend to generate review results as feedback, but intends to allow learners to review their presentation in a self-directed and objective way.
3. Self-Review with Presentation Avatar

3.1 What is Presentation Avatar?

In order to promote self-directed and objective self-review of presentation, we design P-Avatar. The main purpose of P-Avatar is to remove uncomfortable sense brought about in checking out presentation video, which occurs from discrepancy between expected looks/voice and recorded ones. We are currently considering P-Avatar a virtual agent running on computer and an educational robot operating in real world. In this paper, we introduce P-Avatar as a virtual agent.

P-Avatar has two main roles, which are to reproduce presentation learners make, and to produce presentation as their proxy. In presentation reproduction, learners first make preliminary presentation. P-Avatar second records it and captures the slides used, timing of slide change in using their P-document, and their voice/non-verbal behavior. P-Avatar third reproduces the presentation with the slide
information as exactly as possible with the captured data on computer screen. But, the voice is changed with P-Avatar’s one. The learners then check out the presentation reproduced by P-Avatar. In this way, the presentation reproduction allows the learners to objectively self-review their presentation to gain awareness of points to be modified. In order to further enhance such awareness, we are now considering an effective reproduction, in which P-Avatar exaggerates some behaviors different from average ones or some behaviors essential to presentation. Such exaggerated reproduction is out of our scope in this paper.

In proxy presentation, learners do not need to make preliminary presentation, but prepare their P-document, its oral manuscript, timing information for changing the slides in using the P-document, and information about pointing to keywords/sentences as their gesture in advance. P-Avatar uses the prepared data to produce the presentation instead of the learners. They then check out the produced presentation by P-Avatar for presentation improvement. In this way, the proxy presentation allows learners to review presentation expected from the prepared data without making their preliminary presentation. In particular, it would be useful for improving P-document and reconsidering how to use it.

In this paper, we focus on how to promote self-review of presentation made by learners. In the following, let us accordingly demonstrate P-Avatar with the role of presentation reproduction.

### 3.2 P-Avatar System

We have developed a P-Avatar system using Unity Technologies (Unity - Game engine, tools and multiplatform, 2017), in which P-Avatar is implemented as virtual agent on computer. Using Leap Motion (Leap Motion, 2017) and microphone as input devices, the system records and captures presentation learners make as shown in Figure 3 (a). In reproducing the presentation with the captured data, the current system uses the secondary creation of the character "Hatsune Miku" (About HATSUNE MIKU | CRYPTON FUTURE MEDIA, n.d.) of Crypton Future Media, INC as P-Avatar as shown in Figure 3 (a), which is created from the model data "Tda formula Hatsune Miku Append" provided by Tda (Tda formula Hatsune Miku Append Ver1.10 – BowlRoll, 2017). This is created under Piapro Character License (piapro.net, n.d.). The system also integrates MMD 4 Mechanism (Stereoarts Homepage, 2017) created by Nora to make use of the model data on Unity.

Let us demonstrate how to reproduce the presentation with P-Avatar in the following. The system first converts a PPT file to image file, and displays the first slide as image in the user interface. Learners can change the slide into the next/previous one via keyboard. Animation embedded in the slide is turned off. Before presentation, learners can select the use of a pointer.

When learners start self-rehearsal, the system makes P-Avatar invisible, and induces them to pay more attention to their presentation. Instead of P-Avatar, the system displays a marker that indicates the position of the right hand of P-Avatar. If the pointer is used, this marker is not displayed. Their hand gesture is projected as movement of the marker or pointer in real time. The pointer/marker is displayed all the time during presentation. It allows the learners to know to which position/direction their right hand points. When they also change the slide to the next or previous one via keyboard, the system captures the timing. Leap Motion captures their hand gesture that includes pointing to the slide. The microphone records their oral explanation. Their utterance is not projected in real time. The system currently captures only hand gesture, but it seems most important in presentation. That is the main reason we focus on it as presentation gesture.

When the learners finish presentation, the system generates video with the recorded and captured data, in which P-Avatar reproduces the presentation. As for oral explanation, the system transforms the voice quality by adjusting the value of fundamental frequency and formant frequency. The fundamental frequency indicates voice height. The formant frequency also indicates resonance determined by the shape of a vocal tract or an oral cavity, and it characterizes individual voices. When they replay the video, the hand gesture and utterance captured are projected as hand and mouth movement of P-Avatar as shown in Figure 3(b). If the position of the pointer moves to the left of the screen beyond a certain line, P-Avatar faces toward the slide. The presentation slides are displayed and changed in the recorded timing. Figure 3(c) shows an example of presentation reproduction with P-Avatar, in which P-Avatar points to the slide with the pointer to explain.
4. Case Study

4.1 Preparation and Procedure

We had a case study with the purpose of ascertaining whether P-Avatar could promote self-reviewing presentation without uncomfortable sense. Compared to self-review with presentation video, we analyzed how many points to be modified could be found. We also prepared a questionnaire that asked which self-review enhanced awareness of modified points, concentration on review, and uncomfortable sense.

The participants were 9 graduate and undergraduate students in informatics and engineering. We set two conditions: (a) self-review with presentation video (Video condition), and (b) self-review with the P-Avatar system (P-Avatar condition). As within-subject design, each participant conducted self-review twice under these two conditions. The participants were divided into two groups called group V-PA and group PA-V. We randomly assigned 5 participants to group V-PA and 4 participants to group PA-V. Group V-PA first self-reviewed under the Video condition and then self-reviewed under the P-Avatar condition. Group PA-V self-reviewed in the reversed order of the two conditions.

This study included 2 sessions referred as Session I (presentation) and Session II (self-review). Before Session I, all participants were required to prepare their PPT document including the contents...
of their research so that they can make presentation within 15 minutes. In Session I, they were required to carry out preliminary presentation twice with their prepared P-document. They were first required to present with their P-document projected on a screen in front of video camera. The presentation was video-recorded. They were then required to present with the P-Avatar system. They were also informed that they would be required to self-review their presentation two days later.

In Session II conducted after two days, all participants were required to check out their presentation twice with the recorded video and the P-Avatar system, and to find points to be modified. In this session, group V-PA was required to first watch the recorded video and then to use the P-Avatar. Group PA-V was also required to first use the P-Avatar system and then to watch the recorded video. We had an interval of 3 hours between two self-reviews. Before using the P-Avatar system, all participants were given an explanation about how to use it. The use of a pointer in P-Avatar was left up to them. Since P-Avatar was a female virtual character, in this study, the system transformed their recorded voice by raising the fundamental frequency and formant frequency value. In each self-review, we gave all participants a handout of their PPT document. The participants were required to annotate the corresponding slide in the handout with the points to be modified. After Session II, all participants were required to answer a questionnaire, in which they were asked which self-review had more influence on awareness of modified points, concentration on self-review, and uncomfortable sense.

In order to ascertain whether P-Avatar could contribute to finding points to be modified in presentation, we compared the numbers of modified points found in Session II, which we summed up from annotations made in the handouts. We also ascertained the potential of P-Avatar for promoting self-review of presentation from the answers to the questionnaire. The hypotheses we set up in this study were as follows:

- H1: P-Avatar allows finding more points to be modified, and
- H2: P-Avatar provides more awareness of modified points, concentration on self-review, and less uncomfortable sense.

4.2 Results

Table 1 shows the number of points to be modified that were obtained in self-review by the participants. However, this table includes modified points related to slide animation and body gesture in Video condition, which could not be obtained in Self-Avatar condition. It also includes modified points related to operations of P-Avatar, which could not be obtained in Video condition. We accordingly removed the number of such modified points from Table 1 to make Table 2. Although only 3 participants obtained more points to be modified with P-Avatar in Table 2, 7 of 9 participants obtained more or equal modified points with P-Avatar than recorded video in Table 2. In particular, all participants in group PA-V obtained more modified points with P-Avatar than recorded video. In both groups, in addition, the average numbers of modified points per slide in Self-Avatar condition were higher than Video condition in Table 2. These results suggest that hypothesis H1 is supported.

Following the results in Table 2, we further divided the points to be modified into ones for P-document, ones for gesture, and ones for oral as shown in Table 3. Table 4 shows the average numbers of points to be modified for P-document, gesture, and oral obtained in both conditions and both groups. Overall, there were fewer points to be modified for gesture. Except modified points for P-document and oral in group V-PA, the average numbers of modified points per slide obtained in P-Avatar condition were higher than the ones in Video condition. From the result of the one-sided t-test with logarithmic transformation, there was a tendency of significant difference between the average numbers of modified points for P-document in Video condition and P-Avatar condition (t (8) = 0.0675, p < .10).
Table 1: Number of modified points found in self-review.

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant</th>
<th>Number of slides</th>
<th>Modified points</th>
<th>Average points per slide</th>
<th>Modified points</th>
<th>Average points per slide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group V-PA</td>
<td>A</td>
<td>28</td>
<td>35</td>
<td>1.250</td>
<td>24</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>35</td>
<td>37</td>
<td>1.057</td>
<td>28</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>40</td>
<td>17</td>
<td>0.425</td>
<td>9</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>25</td>
<td>12</td>
<td>0.480</td>
<td>14</td>
<td>0.560</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>23</td>
<td>18</td>
<td>0.783</td>
<td>17</td>
<td>0.739</td>
</tr>
<tr>
<td>Group PA-V</td>
<td>F</td>
<td>30</td>
<td>8</td>
<td>0.267</td>
<td>9</td>
<td>0.300</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>32</td>
<td>6</td>
<td>0.188</td>
<td>12</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>40</td>
<td>20</td>
<td>0.500</td>
<td>19</td>
<td>0.475</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>35</td>
<td>16</td>
<td>0.457</td>
<td>13</td>
<td>0.371</td>
</tr>
</tbody>
</table>

Average points per slide in group V-PA: 0.799
Average points per slide in group PA-V: 0.353
Average points per slide in both groups: 0.601

Table 2: Number of modified points except the ones removed.

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant</th>
<th>Number of slides</th>
<th>Modified points</th>
<th>Average points per slide</th>
<th>Modified points</th>
<th>Average points per slide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group V-PA</td>
<td>A</td>
<td>28</td>
<td>24</td>
<td>0.857</td>
<td>24</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>35</td>
<td>30</td>
<td>0.857</td>
<td>28</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>40</td>
<td>13</td>
<td>0.325</td>
<td>9</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>25</td>
<td>8</td>
<td>0.320</td>
<td>14</td>
<td>0.560</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>23</td>
<td>15</td>
<td>0.652</td>
<td>17</td>
<td>0.739</td>
</tr>
<tr>
<td>Group PA-V</td>
<td>F</td>
<td>30</td>
<td>8</td>
<td>0.267</td>
<td>9</td>
<td>0.300</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>32</td>
<td>5</td>
<td>0.156</td>
<td>12</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>40</td>
<td>17</td>
<td>0.425</td>
<td>18</td>
<td>0.450</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>35</td>
<td>12</td>
<td>0.343</td>
<td>13</td>
<td>0.371</td>
</tr>
</tbody>
</table>

Average points per slide in group V-PA: 0.602
Average points per slide in group PA-V: 0.298
Average points per slide in both groups: 0.467

Table 3: Number of points to be modified for P-document, gesture, and oral.

<table>
<thead>
<tr>
<th>Group</th>
<th>Participant</th>
<th>Number Total</th>
<th>P-document Total</th>
<th>Gesture Total</th>
<th>Oral Total</th>
<th>P-Avatar condition Total</th>
<th>Gesture Total</th>
<th>Oral Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group V-PA</td>
<td>A</td>
<td>10</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>11</td>
<td>7</td>
<td>12</td>
<td>20</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Group PA-V</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>
Table 5 next shows the results of the questionnaire, which consisted of four questions. From the results of accurate binomial test for Q1 to Q3, there were significant differences between recorded video and P-Avatar (p = 0.0898, p < .10 for Q2; and p = 0.0020, p < .01 for Q3). The answers to Q1 and Q2 suggest the potential of P-Avatar for promoting awareness of points to be modified and concentration on self-review. The answer to Q3 also suggests that P-Avatar has a potential to remove uncomfortable sense occurring in self-review. These results support H2.

Table 4: Averages of points to be modified per slide.

<table>
<thead>
<tr>
<th></th>
<th>Video condition</th>
<th>P-Avatar condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average points per slide</td>
<td>Average points per slide</td>
</tr>
<tr>
<td>P</td>
<td>0.226</td>
<td>0.077</td>
</tr>
<tr>
<td>G</td>
<td>0.056</td>
<td>0.000</td>
</tr>
<tr>
<td>O</td>
<td>0.150</td>
<td>0.043</td>
</tr>
</tbody>
</table>

† One-sided t-test: p < .10

Table 5: Questionnaire and results.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Which self-review promoted obtaining points to be modified?</td>
<td>Recorded video 3</td>
</tr>
<tr>
<td>Q2. Which was better for concentration on review?</td>
<td>Recorded video 2</td>
</tr>
<tr>
<td>Q3. Which made you feel uncomfortable?</td>
<td>Recorded video 9</td>
</tr>
<tr>
<td>Q4. Was there any inconvenience in using the P-Avatar system?</td>
<td>Yes 7</td>
</tr>
</tbody>
</table>

** p<.01, † p < .10

4.3 Discussion

Table 2 and Table 4 suggest that P-Avatar contributes to finding more points to be modified for presentation particularly for P-document. Table 5 also suggests that P-Avatar enhances awareness of modified points, and concentration on self-review, and decreases uncomfortable sense. From these results, we can say that P-Avatar promotes self-review of presentation.

In this case study, self-review was conducted twice. The second self-review could accordingly obtain more points to be modified than the first review since it is easier to become aware of modified points obtained in the first review. The number of modified points in the first review tends to become lower than the one in the second review. As shown in Table 2, however, all participants in group PA-V obtained more modified points with the P-Avatar system even in the first review. This also suggests that P-Avatar strongly enhance awareness of points to be modified.

In Table 4, there were no significant differences about gesture and oral. As for gesture, five participants did not find any modified points. This suggests that they were able to perform expected gestures sufficiently or they did not regard gestures as modified points. As for oral, three participants pointed out in the questionnaire that the volume of presentation reproduced by P-Avatar was small, and that the sound had a little bit of noise. We need to address these problems as our future work.

As for Q1 and Q2 in the questionnaire shown in Table 5, there were 3 participants who selected recorded video. However, all the participants obtained more points to be modified with P-Avatar. They commented on shortcomings of the P-Avatar system such as no slide animation, difficulty in projecting hand gesture on P-Avatar, low volume of P-Avatar, P-Avatar’s looks, etc., which seem the reasons they
did not select P-Avatar in Q1 and Q2. We need to correct these shortcomings to refine the P-Avatar system, which is our future work.

From the result of Q4 in the questionnaire shown in Table 4, in addition, most of the participants felt inconvenient to use the P-Avatar system. In this case study, there were also points to be modified, which could be found with recorded video and could not be found with the P-Avatar system. The representative points are slide animation, and non-verbal behavior (except hand gestures) such as body movement, gaze, facial expression, etc. In future, we accordingly need to redesign the use of the P-Avatar system, and to address how to reproduce slide animation and how to capture and reproduce the non-verbal behavior to improve the potential of self-review with P-Avatar.

5. Conclusions

In this paper, we have proposed P-Avatar, which promotes self-directed and objective self-review of presentation. The P-Avatar system allows learners to remove uncomfortable sense, which would occurs from their looks and voice on the video. It also allows them to objectively self-review their presentation to gain awareness of points to be modified. The results of the case study with the P-Avatar system suggested that P-Avatar has a potential to promote self-reviewing to improve presentation. In future, we will refine the P-Avatar system so that it can particularly reproduce slide animation and more adequate motion capture. In addition, we will also develop a P-Avatar system, which carries out proxy presentation or exaggerated reproduction of presentation, and which uses educational robot as P-Avatar.

References


Cognitive Investigation of Dynamic Educational Presentation toward Better Utilization of Presentation Characteristics

Yasuhisa OKAZAKI a* & Atsushi YOSHIKAWA b

a Graduate School of Science and Engineering, Saga University, Japan
b Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Japan
* okaz@cc.saga-u.ac.jp

Abstract: Through two cognitive experiments, we investigated fundamental characteristics of presentation methods for teaching materials. We conducted a subjective evaluation by questionnaire and an objective evaluation by comprehension test, as well as gaze analysis with an eye tracker to analyze differences between three cases: when learners were shown a handwriting process such as on a chalkboard, animated slides using presentation software, and only final written results. Results of the experiments confirmed the basic characteristics of each presentation as visual information and provided evidence showing advantages of showing the writing process under certain conditions. We also confirmed an induction effect by voice description as auditory information. These results contribute to the use of various presentation characteristics for maximizing the effect of teaching material presentation.

Keywords: gaze, eye tracker, writing process, chalkboard, slide

1. Introduction

We experimentally examined the characteristics of methods in which teaching materials are presented to learners to identify the benefits of dynamic presentation in focusing on the advantages of chalkboard presentations that present the entire writing process. We believe the results give a criterion when to use a chalkboard and when to use presentation slide.

The use of information and communication technology is expected to promote “understandable lessons,” advancing the use of teaching materials via computers and computer networks (MEXT, 2011; MIC, 2014). Chalkboards and presentation software are widely used to present educational information during lessons. The chalkboard—the traditional presentation method—has advantages with respect to visualization of thinking and flexibility of presentation, because instructors can demonstrate sequential processes and easily add text, annotations, or marks (Brown, 2012; Jenami, 2017). In contrast, presentation software is advantageous with respect to the quantity and quality of presented or reused information, because instructors can quickly present teaching materials including illustrations and videos alongside text, create attractive slides using design functions, and can reuse the materials (El-Ikhan, 2010; UCF - Faculty Center for Teaching and Learning, 2016).

Interactive whiteboards combine the advantages of both slides and traditional chalkboards (TechLearn, 2012). However, they have not yet replaced traditional chalkboards because of the costs for installation and teaching material development. Furthermore, it is necessary to master their operation, they have an inherent potential for breakdowns, and they are smaller than chalkboards.

As an alternative approach, we have been developing a new presentation tool for education that incorporates the characteristics of chalkboards while displaying slides (Hosoki et al., 2011). Realizing a new method of presentation that retains the advantages of chalkboards will require obtaining knowledge about various presentation methods. We believe that handwritten presentation of information, one of the unique characteristics of chalkboards, includes rich educational information because it demonstrates a sort of thought process. Animated slides merely attract attention or mechanically present information step-by-step; they do not reflect the structure of the presented subject or the thinking process. We believe that showing this process through handwriting is meaningful with
respect to comprehension that is different from merely showing completed forms or mechanical stepwise presentation.

As basic data for clarifying the advantages of showing the writing process, we focus on the eye movements of learners and on subjective and objective evaluations of learners through testing (Okazaki et al., 2013; Okazaki, Noguchi & Yoshikawa, 2014). Research has advanced technologies for gaze measurement, and studies on sentence comprehension using gaze tracking devices have been conducted (Duchowski, 2007; Ohno, Mukawa & Yoshikawa, 2002).

We are interested in influences on comprehension that result from showing the writing process. To that end, we investigated differences between dynamic presentations of the writing process, such as writing on a chalkboard, and static presentations of only the final results of writing (Okazaki et al., 2013; Okazaki, Noguchi & Yoshikawa, 2014). Our experimental results showed that dynamic presentation encourages gazes and promotes constructive understanding with step-by-step presentation. We decided that dynamic presentation actively gives information to a receiver of information (learner). We named this fundamental characteristics of dynamic presentations as "information push". Also our experimental results showed that static presentation encourages to interpret given information freely. We decided that static presentation promotes a receiver of information (learner) to extract information. We named this fundamental characteristics of static presentations as "information pull". The "information push" and "information pull" characterize how information is given in dynamic presentation and static presentation. (Okazaki, Noguchi & Yoshikawa, 2014).

This paper considers animated presentation slides as an example of dynamic presentation. We test our proposed idea through comparison with handwriting presentation, and demonstrate fundamental characteristics of each dynamic presentation. We also introduce voice descriptions in educational presentations and examine their effects. Two experiments are described below, a comparison of chalkboard-style and animated slide presentations with no voice, and ones with voice descriptions added.

2. Experiments and Results

2.1 Experiment Method

This section describes the experimental design common to the two experiments. We prepared three presentation patterns with different contents. Three presentation conditions (handwriting, animated slides, and static images) were set for each pattern, comprising nine presentation stimuli with different contents or conditions.

We created handwritten presentation stimuli using a presentation tool currently in development at our laboratory, the Handwriting Presentation Tool (HPT). This tool allows reproduction of writing processes such as writing on a chalkboard (Hosoki et al., 2011). Microsoft PowerPoint was used to make animated slides. Static image presentations showed only the final results of the handwriting.

To investigate and analyze any differences, we used a subjective assessment by questionnaire and an objective assessment by comprehension test (100-point scale) as well as eye trackers for gaze analysis. The Tobii X1 Eye Tracker—a standalone, compact eye tracking unit—was used to track gaze (Tobii, 2015). No distracting sensors or other hardware elements were visible to the user. The system detects the user’s gaze by using both bright and dark pupil tracking techniques. We used Tobii Studio to analyze gaze data. We used gaze plots for dynamic analysis and heat maps for static analysis (Tobii, 2017).

Trial subjects sequentially watched three patterns under one of the presentation conditions. Each presentation was assigned to test subjects according to a combination table known in experimental design as an L9 orthogonal array, where condition combinations are equally arranged.

2.2 Experiment Results

2.2.1 Experiment 1: Comparative Experiment between Handwriting Presentation and Animation Slides
There were 22 participants, all university students in their early twenties. Experiments were performed between February 5–10, 2016. Table 1 shows subjective assessments of the presentation conditions, in total and by condition. Comprehensibility seems to be different between handwriting presentation and animation slides, which are both forms of dynamic presentation. Static images, on the other hand, were the hardest to understand, followed by animated slides and then handwriting presentation. Focusing on the presentation conditions shows interesting results. The evaluation order for handwritten presentation and animated slides is opposite. In the first pattern, handwritten presentations were best evaluated. In the second pattern, however, animation slides were best evaluated, indicating a significant difference in the effectiveness of presentation conditions between the two presentation patterns.

Figure 1 shows average scores and standard errors of the comprehension tests. Patterns 1 and 2 were not easy for participants, while pattern 3 was easy. Consistent with our proposed idea, dynamic presentations are thus useful in these presentation patterns, while static presentations are useful for easy patterns. In addition to confirming our idea, this provides a clue about the difference between handwritten presentations and animated slides.

We conducted gaze analysis by gaze plot and heat map. Gaze plots show the location, order, and time spent looking at locations (Tobii, 2017). We confirmed gaze effects for both animated slides and handwritten presentations. In both cases, gaze points followed the presentation process. One distinctive difference is that gaze could not keep up with the presentations in some cases.

Heat maps show gaze distributions over the stimulus (Tobii, 2017). There were no characteristic differences between the two dynamic presentations (animated slides and handwritten presentation). In contrast, there was a remarkable difference in eye movements between the dynamic

---

Table 1: Difference of presentation method and its subjective evaluation (Experiment 1).

<table>
<thead>
<tr>
<th></th>
<th>Pattern 1</th>
<th></th>
<th>Pattern 2</th>
<th></th>
<th>Pattern 3</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Worst</td>
<td>Best</td>
<td>Worst</td>
<td>Best</td>
<td>Worst</td>
<td></td>
</tr>
<tr>
<td>Handwriting</td>
<td>57%</td>
<td>33%</td>
<td>25%</td>
<td>29%</td>
<td>27%</td>
<td>0%</td>
<td>36% 23%</td>
</tr>
<tr>
<td>Animation Slide</td>
<td>14%</td>
<td>33%</td>
<td>75%</td>
<td>29%</td>
<td>18%</td>
<td>40%</td>
<td>28% 32%</td>
</tr>
<tr>
<td>Static Image</td>
<td>29%</td>
<td>33%</td>
<td>0%</td>
<td>42%</td>
<td>55%</td>
<td>60%</td>
<td>36% 45%</td>
</tr>
</tbody>
</table>

Figure 1. Average score and standard errors of comprehension test (100 point scale) (Experiment 1).
and static presentations (Figure 2). Heat map patterns for static images differed between participants P and G; participant P carefully observed the problem, graphic, and solution (Figure 2(a)), while participant G quickly scanned the problem and solution, indicating that less attention was paid (Figure 2(b)).

Consistent with our previous studies, we confirmed that there are individual differences. We believe this is consistent with the free interpretation of presented information in static presentations.

### 2.2.2 Experiment 2: Effect of Voice Description

There were 20 participants, all university students in their early twenties. Experiments were performed from January 11–20, 2017.

The results were consistent with our previous studies. In both subjective and objective evaluations, handwritten presentations were highly rated for presentations that were difficult to understand, because presentations of the thought process were incomplete in animated slides, which applied a “fade-in” effect to presented slides. In this experiment, we realized again that presenting the thinking process by handwritten presentation is useful.

Table 2 and Figure 3 respectively show subjective assessments of presentation conditions (in total and by condition) and average scores and standard errors of the comprehension tests. The results are mostly consistent with the results of the previous experiment with no voice description. However, the subjective evaluations by the participants and comprehension test results differed; the average score for subjective evaluations was lowest, and that of comprehension tests for static images was highest. This suggests that the difference is a result of voice descriptions.

A similar gaze analysis was also performed in this experiment and showed inductive effects by both dynamic presentation and by voice description. We confirmed that characteristics of free interpretation in the static presentation disappeared as a result of voice description, and that the static image was viewed according to the voice description.

### Table 2: Difference of presentation method and its subjective evaluation (Experiment 2).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Pattern 1</th>
<th>Pattern 2</th>
<th>Pattern 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Worst</td>
<td>Best</td>
<td>Worst</td>
</tr>
<tr>
<td>Handwriting</td>
<td>40%</td>
<td>17%</td>
<td>67%</td>
<td>25%</td>
</tr>
<tr>
<td>Animation</td>
<td>30%</td>
<td>17%</td>
<td>33%</td>
<td>25%</td>
</tr>
<tr>
<td>Slide</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static Image</td>
<td>30%</td>
<td>66%</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>
3. Discussion

In a previous study, we characterized dynamic and static presentations by introducing “information push” and “information pull.”

Through comparative experiments between handwritten presentations and animated slides, this paper demonstrated the same gaze effect for animated slides as seen in handwritten presentations. We also found a difference between the two dynamic presentations (handwritten presentations and animated slides), and demonstrated the possibility that dynamic presentations including the entire writing process are useful for understanding. We believe that this advantage is a result of characteristics of handwritten presentations, which provide gaze points and visualize the thinking process by sequentially presenting objects and their relations.

This experiment showed that handwritten presentations are useful for difficult-to-understand content, because they have a positive influence on sequential constructive understanding.

We also found an inductive effect of voice descriptions on gaze. This effect was observed in both static and dynamic presentations, which we believe may affect comprehension of presented materials.

However, subjective evaluations were lower when voice descriptions were added, especially for easy presentations. We think this is because induction by voice description had a negative influence, due to inadequate speed and timing. This is supported by a questionnaire comment, “I was irritated because the speed of explanation did not match.”

From the above, we conclude that voice descriptions have a significant effect on presentations.

4. Conclusions and Future Work

We investigated the fundamental characteristics of teaching material presentations through two cognitive experiments. Comparisons demonstrated similarities and differences between handwritten presentations and animated slides, and the results were consistent with our previous study. These can be explained by the proposed ideas of “information push” and “information pull.” Our experimental results support the idea that dynamic presentations—which visualize the thinking process—have a positive influence on sequential constructive understanding, and thus are useful for difficult-to-understand presentations.
Further studies are needed to investigate interactions between visual information (shown materials) and auditory information (voice descriptions) when presenting teaching materials. We plan to identify factors that affect understanding through a series of experiments that consider the speed of voice descriptions and their timing during presentations.

Acknowledgements

This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research(C)) Grant Number 16K01117. We would like to thank Prof. Takaaki Sonoda and Ms. Noriko Nakashima at Integrated Center for Educational Research and Development of the Faculty of Education in Saga University for their kind support and cooperation in this research. We also thank to Mr. Kouhei Nishimura, Mr. Kenta Tashiro and all others who cooperated in our experiment.

References

The Journey to Improve Teaching Computer Graphics: A Systematic Review

Thomas SUSELO*, Burkhard C. WÜNSCHE & Andrew LUXTON-REILLY
University of Auckland, New Zealand
*tsus609@aucklanduni.ac.nz

Abstract: Computer graphics is often regarded an exciting and enjoyable subject due to it combining technology, art and creativity. The past few years have seen a rapid evolution of the field with novel consumer-level devices (e.g. head-mounted displays) and media (e.g. 3D videos on YouTube) enabling a much wider section of the population to experience and create 3D content. However, teaching computer graphics can be challenging due to it requiring a diverse range of skills such as mathematics, physics, programming, spatial reasoning, problem solving, and art and design. Several researchers have acknowledged this problem and have attempted to make computer graphics teaching easier and more effective. However, so far no consensus seems to exist about the key problems teachers need to overcome and what concepts and methodologies might help with this. In this paper, we address this issue by conducting a systematic literature review identifying reported challenges, methodologies, and approaches for teaching computer graphics. Our research offers practitioners new insight into computer graphics teaching, which we hope will be useful for curriculum design, developing more effective tools and support for struggling students, and suggesting avenues for future research.

Keywords: computer graphics, systematic review, teaching, learning, teaching methodology

1. Introduction

Computer graphics is a challenging subject to teach and learn (Du and Shu, 2011; Han et al., 2008; He and Zhao, 2012; Peternier et al., 2006; Sueyasu et al., 2010; Zhou et al., 2010) because it involves a wide variety of skills such as mathematics, programming, physics, cognitive psychology, spatial reasoning, problem-solving, human-computer interaction, and art and design.

Over the past decade, many researchers have tried to address this problem and several specialised teaching tools have been developed. For example, Spalter and Tenneson (2006) present an interactive Java-based application called Graphics Teaching Tool (GTT) that allows students to study objects and understand what transformations have been applied to them and what the resulting transformation matrix is. Sueyasu et al. (2010) developed a Simplified Language for Graphics Programming (SLGP) and claimed it improves students' productivity in computer graphics developments significantly. However, no consensus seems to exist about what specific issues make computer graphics difficult to teach, and what teaching methodologies and concepts should be used to address this.

We could not find previous work reviewing research on teaching and learning of computer graphics. In this paper, we will systematically evaluate the literature in order to identify the challenges in teaching computer graphics and identify promising concepts and methodologies for addressing them.

2. Methodology

We use the protocol by Kitchenham et al. (2010) as a framework for identifying relevant literature on teaching and learning of computer graphics and for describing challenges, methodologies, and approaches. Our study goals are captured by the following research questions:

- Q1. What common issues are reported in teaching and learning computer graphics?
- Q2. What approaches in the research articles are used to improve teaching computer graphics?
2.1 Searching Process

For our systematic literature review, we used four well-known digital libraries: ACM Digital Library, IEEExplore, SpringerLink, and ScienceDirect. A detailed search was run on all four databases on a single day (16 April 2017). The following search criteria were employed: (1) the keywords "3D" or "teaching" or "learning" are contained in the metadata (such as title, abstract, keywords); (2) the keywords "graphics" or "computer graphics" are mentioned in the article title. As not all the databases supported advanced searching in the same way, the searching process was adapted where required to obtain equivalent results. The publication date of articles was restricted to the years 2000 and 2017 since around that time consumer-level graphics hardware started to become common and consequently the importance of low-level algorithms (such as rasterization and visibility determination) reduced.

2.2 Data Collection

A total of 622 articles were collected from the four databases with 165 articles being obtained from the full-text collection of the ACM Digital Library and 230 articles from IEEE Explore using the automated filtering configuration. Searching SpringerLink and ScienceDirect resulted in 154 and 73 articles, respectively.

A two stage filtering was used to produce the primary study data set of relevant articles:

1. Exclude papers not about computer graphics teaching (115 articles remaining),
2. Exclude duplicates (45 articles remaining).

In order to analyse the remaining 45 articles, we identified relevant attributes and themes based on key issues discussed in those papers and the data required to answer our research questions. The following attributes/themes were collected for each paper: (1) type of publication, (2) publication venue (journal or conference name), (3) year of publication, (4) identified problems with teaching computer graphics, (5) proposed teaching approaches and methodologies (6), utilised and/or proposed teaching tools (programming language, API, specialised software), and (7) how the effectiveness of the new teaching approach was measured.

3. Results

From the 45 identified papers 15 were published in journals, and 30 were published in conference proceedings. 22 articles were published in venues related to computing education, 14 articles were published in computer graphics venues, and 9 in general Computer Science venues. 10 articles were published between 2000 and 2005, 22 articles between 2006 and 2011, and 13 articles in 2012 or later.

3.1 Issues in Teaching and Learning Computer Graphics

We found that problems related to teaching and learning of computer graphics can be categorised into four key issues as summarised Table 1.

The first issue is insufficient background, especially inadequate skills in mathematics and programming (Cunningham, 2000; Du and Shu, 2011; G IEEE et al., 2008; Han et al., 2008; He and Zhao, 2012; Hitchner and Sowizral, 2000; Hui et al., 2012; Papagiannakis et al., 2014; Santos, 2001; Schweitzer et al., 2011; Talton and Fitzpatrick, 2007). According to Elyan (2012), mathematical algorithms and procedures are important in computer graphics, especially when used to calculate transformations and projections. Programming skills are needed to implement, understand, and experiment with algorithms. Several researchers report that in their studies students were not sufficiently prepared (Haitao et al., 2012; Lowther and Shene, 2000; Papagiannakis et al., 2014).

The second issue is difficulties in understanding geometric concepts such as transformations, projections and 3D modelling (Du and Shu, 2011; Elyan, 2012; Santos, 2001; Schweitzer et al., 2011; Seron et al., 2008; Sung and Shirley, 2004). Sung and Shirley (2004) suggest that these issues arise because students have little visual experience and comprehension of geometric modelling.

The third issue is difficulties in solving logical problems and making the connection between theory, programming, application and final visual effects (Santos, 2001; Schweitzer et al., 2011; Talton
and Fitzpatrick, 2007). Seron et al. (2008) observed that for topics such as global shading and inverse kinematics students struggled most with the technical complexity of the implementations.

The fourth issue is that many students are passive learners and don’t interact much with peers and teachers (Gao and Zhang, 2014; Hui et al., 2012; Li, Huang and Gu, 2009). One suggestion addressing this issue is to use a top-down approach involving group projects. This resulted in increased attention to learning activities, more autonomous learning, and improved teamwork and communication skills.

Table 1: Learning Issues in Computer Graphics.

<table>
<thead>
<tr>
<th>#</th>
<th>Issues</th>
<th>Solutions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insufficient knowledge of mathematics (Glyce et al., 2008; Hui et al., 2012; Zhou et al., 2010) and basic programming (Lowther et al., 2000; Papagiannakis et al., 2014)</td>
<td>Top-down: learn the foundation and the structure of computer graphics knowledge whilst practising the tools (Sung and Shirley, 2004). &lt;br&gt;Hybrid: Using individual programming assignments linked to a set of traditional lecture units with a balance of implementation, algorithms, and mathematical concepts (Lewis, 2012; Guo et al. 2010).</td>
<td>Students’ mathematical concepts increased (Hitchner and Sowizral, 2000) and Students’ programming skills improved (Elyan, 2012)</td>
</tr>
<tr>
<td>2</td>
<td>Difficulties in understanding transformations, projections and 3D geometric modelling (Elyan, 2012).</td>
<td>Top-down: Implement interactive 3D demos to show concepts and techniques in computer graphics (Kadam et al., 2013). &lt;br&gt;Hybrid: Set up a systematic practice-based learning process for students consisting of concept validation, project design, and project training (Schweitzer et al., 2011; Reina et al., 2014).</td>
<td>Improvement in geometrical structures understanding (Sueyasu et al., 2010).</td>
</tr>
<tr>
<td>3</td>
<td>Difficulties in solving logical problems (Hitchner and Sowizral, 2000; Talton and Fitzpatrick, 2007) and making the connection between theory, programming, application and the visual effects (Stephenson and Taube-Schock, 2009).</td>
<td>Top-down: Using a teaching platform with a set of compact applications to demonstrate computer graphics techniques and algorithms (Ganovelli and Corsini, 2009; Spalter and Tenneson, 2006). &lt;br&gt;Hybrid: Pay attention to the relationship between theory and application, and use practical applications to promote communication between students and teacher (Li et al., 2009; Raikar et al., 2015).</td>
<td>Using modern high-level APIs improved students’ mathematics, problem-solving and logical thinking skills (Hitchner and Sowizral, 2000).</td>
</tr>
<tr>
<td>4</td>
<td>Students have become passive learners and don’t interact much with peers and teaching staff (Peternier et al., 2010, Marti et al. 2006).</td>
<td>Top-down: Pay attention to the relationship between theory and application, and use practical applications to promote communication between students and teacher (Li et al., 2009; Raikar et al., 2015). &lt;br&gt;Hybrid &amp; Bottom-up: Motivating and guide students to be active learners. (Yang and Sanver, 2002; Taxén, 2004; Zhao et al., 2005).</td>
<td>Students’ attention in their learning activities increased, and students became autonomous learners (Gousie, 2000). Teamwork and communication skills improved (Tori et al., 2006).</td>
</tr>
</tbody>
</table>

3.2 Teaching Approach and Methodology

We identified three common approaches of teaching computer graphics. Of the 45 reviewed papers 34 described a top-down approach, seven a hybrid approach, and four a bottom-up approach.
The bottom-up approach is regarded as most traditional (Sung and Shirley, 2004) and seems to be favoured in textbooks for teaching computer graphics and is also most popular based on our personal teaching experience. The approach presents first foundations such as transformations and rendering of simple objects, before more complex topics are introduced (Taxén, 2004; Bouvier, 2002; Cunningham, 2000; Lowther and Shene, 2000; Shanshan et al., 2008; Sung and Shirley, 2004).

The top-down approach starts with a moderately complex problem or case study, e.g. a simple game, and then breaks it down into simpler problems (functional modules) (Shirley et al., 2015; Sung and Shirley, 2004, Talton and Fitzpatrick, 2007). This can help students absorbing the foundations and structure of graphics applications while practising (visible rather than mathematical based) application-level understanding and skills (Gousie, 2000; Santos, 2001; Seron et al., 2008; Song et al., 2009; Sueyasu et al., 2010; Sung and Shirley, 2004; Tori et al., 2006; Yang and Sanver, 2002). The top-down approach often involves using high-level tools and can promote self-learning and increase students’ motivation and knowledge of fundamental computer graphics concepts (Nishino et al., 2011, Song et al., 2009, Sueyasu et al., 2010). Using high-level development tools, such as game engines, can enable students with insufficient mathematics skills to understand computer graphics concepts and produce attractive results (Elyan, 2012; Nishino et al., 2011; Raikar et al., 2015; Shanshan et al., 2008; Sueyasu et al., 2010; Tori et al., 2006).

Several researchers have combined the top-down and bottom-up approach into a hybrid approach (Glvez et al., 2008; He and Zhao, 2012; Hitchner and Sowizral, 2000; Hui et al., 2012; Schweitzer et al., 2011; Tori et al., 2006). The motivation is to support the learning of practical skills while simultaneously improving the knowledge base. Teachers can combine theory with graphics programming and use of graphics software to foster students’ abilities in solving practical problems (Andújar and Vázquez, 2006; Reina et al., 2014; He and Zhao, 2012; Hui et al., 2012; Schweitzer et al., 2011; Tori et al., 2006).

4. Discussion

Our research identified four key issues encountered when teaching computer graphics. Three of the issues were related to insufficient skills and/or a difficulty of applying skills related to mathematics, programming, and spatial reasoning. The fourth key issue was that students have become passive learners. The results suggest that more care must be taken in defining appropriate prerequisites for entering a computer graphics course, or that more emphasis needs to be placed on teaching fundamental skills not necessarily directly related to computer graphics.

Our review suggests that a top-down or hybrid approach might be able to address some of these issues by demonstrating the importance and usefulness of fundamental concepts, applying them in a practical context, and giving students an increased motivation to study them (e.g. in order to create a game or other fun application). However, it is unclear whether such a top-down approach will result in the same breadth of knowledge of fundamental concepts as a bottom-up approach, and so far no quantitative comparison of these approaches exists.

Another possibility might be to exploit that students have different skills. We haven’t found any research paper suggesting this approach, but we are aware of many tertiary training institutes, e.g. animation and design schools, combining students with an arts/design and programming background in the same class, but offering different assessment tasks to them.

It is somehow surprising that the vast majority of papers suggest using a top-down or hybrid approach, whereas the traditional bottom-up approach still seems to be most common in practice. One issue might be the fact that a top-down approach is often project-based, e.g. developing a simple game (Ganovelli and Corsini, 2009). This is likely to be attractive to students, but requires a large amount of supervision and might be difficult to implement for very large classes or instances where students’ abilities differ substantially. None of the papers we reviewed was about teaching a large class (say, 200 or more students).

There seems to be a general consensus that increased interaction, e.g. via supervised labs, group work or peer work is useful. This corresponds to suggestions made in the computing and general education literature, e.g. the “blended learning” and “flipped classroom” concepts (Lage et al., 2000).
5. Conclusion and Future Works

Our research had identified four key issues making teaching and learning of computer graphics difficult. We discussed these issues and offered suggestions to address them. The most common solution offered in the literature is to employ a top-down or hybrid approach and promote more practical work and more interactions. While these suggestions are common in other educational fields as well, we believe they are particularly important in computer graphics due to the wide variety of skills involved and the fact that its output are images and models. Students can easily relate to visual output and often have an intuitive understanding of it, as compared to, say, the underlying mathematical and physical concepts such as transformation matrices and illumination equations. In future work, we would like to analyse tools and technologies for teaching computer graphics and investigate more formally what type of skills are predictors for success in computer graphics.

References


Designing a "Three Rings" Theory Framework for Electronic Schoolbag

Baoyuan Yin, Fati Wu*, Shihua Huang & Shan Jia
School of educational technology, Beijing normal university, China
*wft@bnu.edu.cn

Abstract: This paper designs a "Three Rings" theory framework for electronic schoolbag. This model is based on three important principles of designing e-schoolbag. The first principle is that e-schoolbag should be applied to classroom through providing the service for teachers. The second principle is that e-schoolbag should integrate the key elements of the teaching system into a whole through putting the activities in the center. The third principle is that e-schoolbag should be applied to the smart learning environments. "Three Rings" theory framework is composed of inner ring, middle ring and outer ring. The inner ring is the sequence of activities. The middle ring is the teaching system composed of social network, resource network, tool network. The outer ring is the source of adaptivity driven by data which have five functions: perception, processing, analysis, evaluation, recommendation.

Keywords: electronic schoolbag; principle; model; theory framework

1. Background

E-schoolbag is an informational portable terminal equipment for teaching and learning. It always use in K-12 education. It provides communication and exchange facilities particularly well adapted to education. It helps the users (learners, teachers, parents) to access resources and services through a single interface, their virtual desktop, available at any time, whatever their hardware, software, network infrastructure, or geographical location (Chabert et al., 2006). Some research indicate the central role of the electronic schoolbag (or e-schoolbag) and its textbook in traditional classroom practice (Li, Chen, & Kulm, 2009). E-schoolbag have been teaching experiment for many years in China, but the educational effectiveness is still controversial. Why is the schoolbag difficult to applied in the classroom effectively? This is an important question that we have to consider in designing an e-schoolbag platform.

There are two important factors to promote e-schoolbag using in classroom in Chinese K-12 education. One is the powerful promotion by the government. The other is the choice of schools and teachers. The former facilitates the usage of e-schoolbag in schools, providing the opportunity for the deep integration of technology and education. The latter is related to the real users of the e-schoolbag and the practitioners of the deep integration of technology and education at the same time. Especially the teachers who are the most important gatekeepers of using technology in the classroom.

2. The Important Principles of Designing e-Schoolbag

2.1 Applying e-Schoolbag to Classroom through Providing the Service for Teachers

Teacher is the key point of in the process of deep integration of information technology and education. He/she decides which technology can enter his/her classroom and how to use it. The teacher is the real gatekeeper of using technology in the classroom. So when designing e-schoolbag, we have to think about how to let teachers willing to use e-schoolbag in his (her) classroom. K-12 teachers have a lot of complicated teaching and management work every day. If using e-schoolbag will increase their work burden, it is difficult to make teachers willing to use e-schoolbag. Therefore, one of the ideas of designing e-schoolbag is avoiding complex technology, simplifying the operation of technology and
making use of technology more simply and friendly, so that teachers can concentrate on their own teaching. Then teachers’ work will become more easily and effectively.

2.2 Integrating Various Elements of the System by Virtue of Activity-centered Teaching

Classroom is the key area of deep integration of information technology and education. The only measure that information technology has a revolutionary impact on education is the changing of teaching structure in traditional classroom. This is reflected by the changing of the status and role of the teaching system elements (Kekang, 2014). There are four essential elements in a teaching system: teachers, students, contents, and media. In the classroom using the e-schoolbag, teachers and students are humans, the content can be expanded into resources, the media is the tools of e-schoolbag. Therefore, through putting the activities in the center, e-schoolbag can integrate the key elements of the teaching system into a whole at the same time, teachers can organize the activities which integrate humans, resources and media into a whole.

2.3 Applying to Smart Learning Environments

Nowadays the new technologies such as cloud computing, large data, pervasive computing and social networks are emerging, which provide a realistic foundation and technical support for the construction of smart learning environments (Kang & Dossick, 2015). Smart learning environments are defined as physical environments that are enriched with digital, context-aware and adaptive devices, to promote better and faster learning (Koper, 2014). E-schoolbag can be applied not only in the traditional learning environments but also in the smart learning environments. E-schoolbag platform used in smart learning environments such as smart classrooms can provide corresponding services for teachers and students according to the characteristics of learners as well as process, analyze, evaluate the information. Therefore, e-schoolbag should have five functions: perception, processing, analysis, evaluation, recommendation.


The theoretical framework of the e-schoolbag is composed of inner ring, middle ring and outer ring. The inner ring of the model is the sequence of activities. The middle ring is the teaching system composed of human, resource and tool. The outer ring is a data-driven intelligent system. In the middle, human, resource and tool are represented by social network, resource network, tool network respectively. They communicate with each other and constitute the learning community of network. The outer ring provides support to the learning community of network. In a smart learning environment, through organized activities teachers integrate various elements of the teaching system (human, resource, tool) supported by the source of adaptivity (Fig. 1).

![Fig. 1. "Three Rings" theory framework.](image1)
![Fig. 2. Design the sequence of activities.](image2)
3.1 The Inner Ring: Sequence of Activities

Learning activities is the sum of behaviors of teachers and students to achieve a specific learning objective. Different learning activities constitute different learning process or learning models. This reflects the different teaching strategies which point to the different teaching objectives. According to the theory of activity, the design of e-schoolbag can be regarded as a linear or non-linear sequence of activities with a specific purpose. A complete process of teaching or learning must include a number of learning activities. An activity may contains multiple tasks or only one task (Wu Fati, 2003). The goal determines the task in the goal-oriented learning activities. According to the characteristics of the task, we make the appropriate arrangements for the sequence of the activities and then refine the design of each learning activities gradually.

3.1.1 Pre-class

Before the class, the teacher carries out instructional design through schedule the sequence of activities(Fig.2). There are common activities used by teachers and students in the platform of e-schoolbag. Each kind of activity in the platform are represented by a special icon. The platform distinguishes special activities according to the disciplines. Such as in physics teaching, the same activities are practice, examination, discussion, display, explain, sharing, evaluation, voting and the special activities include experiment, etc. When a teacher wants to design activities, he or she only needs to drag the icons to the process line, then the system will pop up the corresponding link (such as ppt, video, animation, etc.). The teacher complete the content settings of the activities by clicking on the mouse. When the teacher complete setting all the activities, the sequence of scheduled activities is formed. Then the teacher can upload the resources and sent the materials to students.

3.1.2 In Class

In the class, the teacher carries out teaching according to the sequence of the designed activities (Fig. 2). The scheduled sequence of activities can be changed and deleted at any time and the new sequence of activities can be generated at any time. The platform can diagnosis the problems of students based on the data of learning process and output diagnostic results. These can provide decision-making and evaluation for teachers and students. Homework or test can examine the learning results in the stage. Students can seek help from experts, teachers and companions in the social network. The platform can intelligently evaluate the student's homework or test. The results of the evaluation will be recorded in the student's electronic file.

3.1.3 After Class

The activities after class are the process of digesting, reshaping and expanding knowledge. The process is to expand, extend and secure individual knowledge network based on the results of early diagnosis. The student determines his own learning path based on the diagnostic result. The platform will collect time, space, scene data of the student, then put them into the student model. In the next step the platform recommends the individualized resources to the student based on the learning style, learning needs, learning ability and other personal information. At the same time, the platform also recommends social network to the students. Students can approve and add them into their own network. Then the individualized learning circle is formed that can support the exchange and collaboration between students.

3.2 The Middle Ring: the Learning Community of Network

At the beginning the “Community” was introduced into the subject area as the concept of sociology. Dewey introduced the "community" into education and put forward the concept of “learning community”. Learning community of network is an organization of learning on online. We should not only consider the cultural and situational aspects of the learning community of network, but also think them as a system which can achieve the whole function. Human, resource and tool are the basic
elements of the teaching system and each of them can be seen as a network respectively (social network, resource network, tool network). They together constitute the learning community of network. The learning community of network not only represents the relationship of them, but also expects them should emerge the characteristics as a whole.

Connection is the foundation for the learning community of network. From the perspective of learning theory of connectivism (Siemens & Conole, 2011), the network is composed of numerous nodes. The closer connection between nodes, the stronger the adaptability of system. The seamless connection of social network, resource network and tool network can make learning occur at any time. The process of learning is starting from learners’ internal network, connecting the external nodes, extracting reasonable ingredients, opening up the internal cognitive neural network, reorganizing the concept of the network, and ultimately the personalized learning network is formed.

3.2.1 Social Network

Social network is composed of students, teachers, administrators, experts, etc. In a social network, each person of a learning community is thought as a node and the relationship between people is thought as an edge. Nodes and edges make up the map of the network. In the learning process, nodes can communicate, collaborate and evaluate at any time. Individual node can connect with existing nodes or external nodes that can be traced to get feedback and guidance.

Some important nodes (such as teachers, experts, etc.) in the social network can help student to expand the breadth and depth of knowledge. Therefore, the platform of e-schoolbag should automatically help student trace the important nodes and recommend information such as e-mail, blog, website of experts. The interaction between the resource network and the tool network is facilitated to connect the social network as well. When a single node (student A) is preferred to trace a resource node and a tool node, student A will find student B who have the same preference. They may form a small learning community with others who have the same learning needs and preferences. The diversity of small learning communities is conducive to the ecological development of social networks.

3.2.2 Resource Network

The resource network is a collection of resources related to learning contents, learning outcomes, auxiliary resources and so on. Learning contents include micro learning resource, PPT, animation and so on; learning results include learning log, learning reflection, learning evaluation, learning products and so on; auxiliary resources include item banks, materials and so on.

The resource network integrates the learner model and the resource matching. On the one hand, the system provides the personalized recommendation through the technology of situational awareness, data mining and intelligent aggregation. The high-quality resources will be recommended to teachers and students preferentially. On the other hand, through the continuously connecting and creating new resource nodes, a unique personal resource network can be formed. The diversification, selectivity, availability and timely accessibility of the nodes will change the distribution of information resources and will achieve the reconstruction of the educational relationship in essence.

3.2.3 Tool Network

The tool network is a collection of tools. It includes preparing lessons tool, choosing questions tool and evaluation tool for teachers. It also includes homework tool, wrong answer tool for students. It includes searching tool, communication tools and collaboration tool for all. Tools can be matched and recommended on the basis of the task and the student's needs. The combined using of these tools can support the production, processing, sharing of knowledge. It can also support exchanging and interaction of information between people.

In the learning process, students accomplish task of learning through connecting resources, tools and humans. At the same time, each network continues to eliminate the weak nodes and generate new nodes, so that network will continue to evolve and ensure learning community is optimized. In short, social networks, resource network and tools network come into an entirety who have mutual effects. If one party has a problem, teaching activities cannot be carried out smoothly.
3.3 The Outer Ring: the Source of Adaptivity

The source of adaptivity in outer ring is data-driven composed of perception, processing, analysis, evaluation and recommendation. They provide support to the network learning community by offering information collection, processing and services.

3.3.1 Perception

Perception is that the platform of e-schoolbag can get students’ all kinds of data in the learning process. The platform percepts students' physiological information, environmental information and relationship information through the mobile phone, PAD, wearable equipment, sensors and other sensory. Physiological information includes breathing, blood pressure, pulse, heart rate, body temperature, etc.; environmental information includes time, location, scene, temperature, speed, etc.; relationship information includes social networking partners, teachers, experts, online status, etc.. This multi-source, multi-scene and multi-modal data aggregation makes the diagnosis, prediction and intervention becoming more accurate.

3.3.2 Processing

Processing is that the platform of e-schoolbag can convert the unstructured data to the structured data through the individualized student model. Building an individualized student model is important to intelligent processing. At present, the common individualized student models include coverage model, deviation model, cognitive model and Bayesian model.

The coverage model is based on the comparison between performance of students and expert knowledge to judge the students’ situation. The deviation model constructs the student's behavior model based on expressing the students' misconceptions as the deviation of the expert knowledge. The cognitive model quantifies and measures the student's cognitive ability to reflect the student's knowledge level, cognitive ability and psychological factors. The Bayesian model constructs model by determining the network structure and conditional probability distribution.

3.3.3 Analysis

Analysis is that the platform of e-schoolbag can use technique of learning analytics to analyze the data of learning process, then provide individual learning diagnosis, learning decision, precision recommendation, multiple evaluation and other learning services. The quantification of the learning process makes it possible to analyze the data of the whole process. The learning process data can describe and explain the existing phenomena, early warn and intervene the learning process and predict future learning. With the data collection gradually breaking the technical constraints, the data collection in classroom will not be limited in the questionnaire survey, artificial observation and other methods, but also record students’ facial expressions, eye movement and other data through the perception of equipment. It will greatly enrich the type of data. It will become a reality to build a learner-centered holographic digital model.

3.3.4 Evaluation

Evaluation is that the platform of e-schoolbag can evaluate, predict students’ academic achievement and output the results. Visualized dashboards can help teachers and students understand the process and results of learning activities intuitively such as participation, goal completion, learning ability, knowledge map, etc. This information can help students to construct self-awareness, self-reflection and meaning (Siemens, G., 2013). At present, there are three methods for evaluation and prediction. The first is the use of statistical methods to analysis and predict the students’ behaviors and academic achievements. The second is to use the labeled behavioral samples to establish a multi-feature fusion classification model, mining the data of behaviors and identify unknown samples. The third is to model
the dynamic evolution of learners' behavior in time series and predict the results of learning dynamically.

3.3.5 Recommendation

Intelligent recommendation is that the platform of e-schoolbag can quickly and accurately choose the most appropriate information and provide personalized information services to students. The algorithm is a key part of the design, which determines the quality of service. At present, the common recommendation algorithm is based on the recommendation algorithm of association rules, content-based algorithm, collaborative algorithm based on collaborative filtering and hybrid recommendation algorithm. The hybrid recommendation algorithm is better than a single recommended algorithm, because it can avoid the weakness of other recommended algorithms.

4 Conclusion and Perspectives

In this article, we have presented a “Three Rings” theory framework for Electronic Schoolbag. This theory framework gives an example for designing an e-schoolbag system or other similar adopt systems. This theory framework inspires the designers that the educational theory, the applicability of technology, the local policy and market demand are important factors when they design an adopt system.

This theory framework is the theoretical model of Standard Project of Chinese Education Informatization. The standardization of educational information is of great significance. From the macroscopic point of view, it is of great significance to formulate a unified standard that is conducive to the steady development of the education informatization. From the micro point of view, the unified standard can provide help to the enterprises who develop the products for education informatization. Unified standards of application is the guarantee of market and can avoid repetitive inputs and low level development. Finally, we should research the practical application of the "three-ring" model, then test and modify it for the next step.

Acknowledgements

This study is supported by Program for Chinese Education Science "Twelfth Five-Year Plan" of Ministry of Education Key Project (Project No. DCA140230) and University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (Project No. 120400356)

References

Interaction between Standardisation and Research in Drafting an International Specification on Learning Analytics

Tore Hoel & Weiqin Chen
Oslo and Akershus University of Applied Sciences, Norway
tore.hoel@hioa.no

Abstract: Interoperability standards are key enablers for widespread adoption of learning technologies, e.g., new data-driven analytics of learning, an application domain explored in a case study presented in this paper. Standards-making is a design practice that relies on input from research and end-users, involving experts that represent diverse stakeholders spread all over the globe. However, the standards-setting culture and formal rules are sometimes at odds with the culture and practice of research. Based on previous research identifying lack of openness and transparency, and a suboptimal interaction with academic research as issues that could explain lack of success in a European setting, this paper studies how an ongoing international standards project on privacy and data protection policies for learning analytics has interacted with an international academic research community. The results of this study show that establishing feedback loops between standardisation, research, and development is essential in order to produce results.

Keywords: Standardisation, Design Practice, Learning Analytics, Interoperability, Learning Analytics Systems Design, Privacy, Data Protection

1. Introduction

Lack of standards is often mentioned as a bottleneck when discussing how to mainstream new learning technologies and innovative educational practices (Ecke & Pinto, 2008; Vries & Choi, 2010). Per definition bottlenecks should be removed; therefore, one would expect that a lot of resources and manpower were allocated to standards work. This is not always the case; there is a huge gap between the acclaimed need for standards and the actual participation and activity in standards fora.

This paper focusses on a particular challenge of the standards-setting process, namely how interaction between the research and the standardisation communities could be facilitated in order to solicit necessary requirements and ideas for design. The project used as a case in this paper is under the auspices of the sub-committee 36 of the Joint Technical Committee 1 of ISO/IEC (SC36), which is developing standards for learning analytics interoperability (LAI). Learning analytics (LA) is a new domain of applications and practices driven by the easy access to data provided by mobile devices and an increasing number of sensors. The aim is to achieve actionable insights from data derived from the full spectrum of learning and teaching activities. By sourcing analytics with data from both within and outside of formal institutional settings, LA has the potential to boost system integration in learning, education and training (LET), bringing both institutions and vendors together.

In the following we will establish a backdrop on which the case study will be projected. From this background, questions arise related to how to optimise the standards development process by interfacing with academic research. This will be explored in depth in the foreground study.

2. Backdrop: ICT Standardisation for Learning, Education and Training

Standards' dynamics (Egyedi & Sherif, 2008), especially in the field of anticipatory standardisation, imply that specifications and technologies co-evolve, something that requires a well-coordinated interaction between the standards community and the R&D community.
The ways the two communities organise their work, however, are different, and that to a degree that potentially leads to conflicts. The research community is used to confidentiality and strict governance of IPR. Nevertheless, general design ideas and opinions are shared openly. When the standardisation hat is put on, a researcher may experience a different culture, where the norm is secrecy and uncertainty whether non-controversial information may be shared.

Hoel (2014) concluded that the document-for-profit model of formal standards bodies drives a wedge between the standards community and the research community. When the sustainability of the standardisation system rests on the sale of documents open distribution of drafts for input and comments becomes a threat to the standards organisation. In 2014 the European Committee for Standardization (CEN) Workshop on Learning Technologies was disbanded by CEN Technical Board after years of conflict about working process and procedures with the workshop's own experts. The outputs of the Workshop were the basis for standards development in the CEN Technical Committee 353, and when the Workshop died, the TC went to sleep the year after (Hoel, 2014).

3. Foreground: in Search of Input and Comments on Privacy and Data Protection

Reflecting on the background research we see that lack of openness and transparency could be operationalised as lack or insufficient exchange between stakeholders that play different roles in standards-setting and use of the outputs, i.e., (1) the research community, (2) the standards practitioners, and (3) the users of standards. Low output and even low technical quality (Hoel & Mason, 2011) could be attributed to insufficient input from research and development, and insufficient testing and feedback from the implementers of standards. How could this process be improved when embarking upon a new project within the domain of LAI?

Standardisation is a design practice, and therefore it would be worthwhile to look at design science research methodology to learn more about how to design processes for knowledgeable outputs. According to Gregor and Hevner (2013, p. 345) Design Science Research (DSR) activities are positioned in one of four quadrants in the cross-section of application domain maturity and solution maturity. The field of LA (and therefore LAI) is quite immature, both in terms of conceptual understanding and access to applications. Therefore, the solution maturity is low, which positions the design activities as invention of new solutions for new problems, contributing to knowledge creation and exploration of research opportunities.

While DSR contributes to both descriptive and prescriptive knowledge creation (Gregor & Hevner, 2013, p. 344), the main objective of standardisation will always be to harness prescriptive knowledge. Action Design Research, a near-standing field to DSR, is defined by Sein, Henfridsson, Purao, Rossi, and Lindgren (2011, p 40) as “a research method for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting”. Figure 1 is an adaptation of Sein et al.’s generic schema for IT-dominant Building of the IT artefact, Intervention in the organisation, and Evaluation (BIE) (ibid., p. 42).
The model in Figure 1 assumes that a project initiated in a standards group actively seeks input from research, testing out the developed draft concepts and design ideas with the implementers community through several iterations before finally agreeing to go for a final design, e.g., setting a standard. The output of the BIE process is a contribution to the knowledge base. In Figure 1 we have included the output from the background research, i.e., the importance of academic input, openness and transparency, and open standards for testing and implementation in the adoption community. With the ideal process described in the adapted BIE model in mind, below, we will report the foreground research on standardisation processes observed in a particular project, which addresses challenges of privacy and data protection related to LA.

3.1 Launching a New ISO/IEC TR 20748-4 Project

International standardisation is done according to directives regulating how to establish projects, develop drafts, building consensus, etc. Formal standardisation on national, regional and international level tend to follow similar rules as found in the ISO directives (ISO/IEC, 2016). Technical work is done in technical committees or working groups. Quality assurance and publishing is done by the standard body's management organisation, which is represented in the standards group by a secretary that makes sure the document centric process is followed by carefully archiving written records of progression of work.

Technical work should done be according to the directives, both in spirit and letter; however, sometimes the two are not easily consolidated. Standards experts want to find solutions to wicked technical problems; the standards bureaucracy wants adherence to rules. For example, if appointment by a national body is necessary to take a seat at a working group (WG) table, one cannot just invite a domain expert out of the blue because of possible valuable input. Or, maybe there are ways to combine innovative specification with strict formality?

In the following case study, we present the SC36 project “20748-4”, based on participatory observation account of how the lead editor of 20748-4 has experienced the drafting of the technical report.

Preliminaries: The drafting of the reference model of LAI (ISO/IEC TR 20748-1, 2016) – starting 2015 – identified privacy and data protection policies as a cross-cutting concern affecting all LA processes. It had been a struggle to make sure these issues were represented in the model, as privacy had not yet surfaced as an important issue related to LA in some constituencies, and in some standards-setting consortia privacy was beyond the scope of LA systems (Hoel & Chen, 2016).

Working group context: WG8, the working group in question, is the latest WG to be established in SC36, with participation from a wide range of countries, e.g., Australia, Canada, China, France, Japan, Korea, Norway, and UK. Traditionally, editorial roles have been allocated between participants with representativeness in mind, even if the number of active editors has not always matched the nominal number. For the new project, editors from Norway, Canada, Korea and Japan were approved (joined by a second Korean editor at a later stage).

Drafting process: The key to a good drafting process is a well-defined scope (Hoel & Mason, 2012; 2011). The scope of 20748-4 was to specify attributes and requirements for privacy and data protection with the purpose to inform design of LA systems development and LA practices. In delivering on this scope, it is a challenge to solicit requirements and other input, knowing that formal standardisation of this type does have a major problem in engaging with stakeholders that walk the talk (Hoel, 2013).

Another challenge is related to the drafting and consensus process itself. The process is document centric, with emphasis on version tracking and storing in a dedicated repository. Once the document is circulated as a working draft at the preparatory stage, experience from participation in SC36 working groups shows that it is very difficult to suggest restructuring of the text or adding new perspectives. When formal commenting is initiated – with each national body entering comments into a spreadsheet, detailing the issue related to specific text fragments, and suggesting replacement text – the drafting changes mode and takes the form of wordsmithing. Therefore, it is essential to present a draft that is as coherent and finished as possible, before it is being discussed in the working group (and even in the editorial group when it consists of several persons). In some projects, this challenge is addressed by initiating a study period, which could end up with ideas for a draft text. However, in the case of 20748-4 the editorial group was supposed to develop the first working draft from scratch.
Standards drafting as part of research: Even though there was a formal call for contributions, the lead editor of 20748-4 knew that the necessary input solicitation and testing of ideas had to take place outside the standardisation process as such. With the European debacle of the CEN working group fresh in mind (see section 2), it was clear that to push the envelope one had to do a balancing act making sure the formal statues of ISO were observed while brainstorming design ideas. In practical terms that meant keeping a paper trail, feeding the document registry and organise meetings. Research papers that discussed and tested ideas and perspectives were contributed as experts' contributions; WG8 meetings were co-located with academic conferences and meetings; and academic workshops were organised to discuss privacy and data protection issues, engaging the national experts that later would have formal roles in the standard-setting group.

Consensus process: Without knowing the result of the project under study, we can only report on the processes that we have observed so far. By establishing conduits between a research community with an ongoing conversation about issues of ethics and privacy for LA, and the standards community we have created an influx of viewpoints and perspectives that also is reflected in the draft project document. When co-editors step up to representing national positions we will see how draft text will be evaluated against different conceptions that could take the document in different directions.

4. Attempting Innovation while Adhering to the Rules

Projecting the 20748-4 case onto the idealised model of standards-setting coming out of our background research (Figure 1), we see that there is only a partial fit. The interaction between academic researchers and the standard group participants was established; however, the interaction with the users of standards seems to be missing. One might say that user perspectives were communicated through workshops organised as part of academic conferences. But there is no systematic testing of design concepts that are part of the 20748-4 project. This is a weakness, however, that is inherent in anticipatory standardisation, where there is no clearly defined need when projects are initiated, and where the stakeholders are busy inventing new technologies, with no time for applying standards to level the playing field (Baskin, Krechmer, & Sherif, 1998; Jakobs, 2003; Umapathy, Purao, & Bagby, 2011). How the technical report on privacy and data protection policies for LAI will be received by vendors and educational stakeholders will only be known after publication. However, it is clear from the start that standards of this nature need to go through several development cycles to be able to serve its purpose.

Another observation comparing the case with the model in Figure 1 is that, in practice, there is an overlap between the roles of academic researchers and standards practitioners. In Action Design Research, teams are built where researchers work together with practitioners to design and test artefacts. In research on how Research and Development (R&D) interact with standardisation one has focussed on how the different institutional contexts interact, and which barriers there are for effective knowledge and technology transfer (Interest, 2007).

In the case we have reported, the role as researcher and the role as standards practitioner are often maintained by the same person. However, the acting out of the particular role is heavily influenced by the setting. In SC36, some participants fill roles as professors at national universities, and when observed in their own cultural context they act, as expected, very strongly and vociferously. In the setting of an international standards meeting, however, many of the same persons are hardly uttering a word and are very reluctant to expose their obvious mastery of the subjects in question. In order to establish the necessary basis for any design to take place, this pattern of acting out established roles needs to be broken. The work culture and directives of the formal standards organisation serve, as we have shown, as a considerable barrier against taking on multiple roles, switching between representing one's country or a stakeholder group, and entering a more open brainstorming and creative role. Therefore, in standards-making of the type described in this paper, there is a need to establish a repertoire of instruments to be used to soften the barriers against crossing role barriers.

What instruments do standards experts have in their toolbox to increase the knowledge base, on which anticipatory specification work builds? Are the rules intended to protect intellectual property and the standardisation organisation's business interests barriers to knowledge exchange?

In the case of subcommittee like SC36, the influence of the central Technical Management Board (TMB) is mostly felt when projects are marked red because the deadlines are exceeded. How information is exchanged and the experts communicate are not interfered with from ISO TMB,
providing the usual paper trail is followed and the committee as such is not under special observation because of mismanagement or conflicts. If the experts want to do expansive knowledge seeking and exchange, not much could prevent them from doing so. The barriers are mostly cultural. Formalities are invoked only if there are disagreements, as long as the minimum level ISO document management process is followed.

5. Discussion – Identifying Research Gaps

Standardisation work typically involves conceptual, technical, and political activities that together are focused on achieving consensus among a group of stakeholders. The outcome – a standard – is essentially just a document that represents a stable reference point and sometimes includes detailed technical specifications. How this document is viewed, however, both by the stakeholders and the standard-makers may differ considerably (Ecke et al., 2008). In some countries, standards are seen as vehicles for execution of national or regional government policies. China may here serve as a case in point. While in other countries more driven by market economy, like in Europe and USA, standards are recommendations that is up to the market to embrace. We would therefore claim that to understand the process and outcome of a particular standardisation process, one need to understand the national policy context of the national experts taking part in the project. This is an area where little research is done till now.

Standardisation is also a design activity that has much in common with innovation processes, both in the technical, organisational and political fields. In this paper we have pointed to the importance of relating to methods and approached from the academic research field, where for example openness and transparency play important roles in promoting innovation. Interestingly, when the European Commission in 2017 published a new and updated version of the European Interoperability Framework (EIF) these principles got a prominent role (EC, 2017). In the new version openness is an underlying principles that is defined in terms of a preference for open data (Recommendation 2), open source (Rec. 3), and open specifications (Rec. 4). The new version of EIF also underlines the principle of transparency. In the EIF context, transparency refers to enabling visibility (“allowing other public administrations, citizens and businesses to view and understand administrative rules, processes, data, services and decision-making”); ensuring availability of interfaces with internal information systems; and securing the right to the protection of personal data. Under which conditions in a standardisation setting will innovation thrive, and what roles do the academic research principles like openness and transparency play for the process of standards-making and quality of specifications? This is another under-researched field we have identified in this paper.

Standardisation processes are also about group dynamics, often in a multi-cultural setting. It is important to understand how particular groups deal with the different processes of standards-making, described in Fomin, Keil, and Lyytinen (2003) as Design, Sense-making, and Negotiation. Hoel and Pawlowski (2012) expanded on that model and constructed the new concept of Key Knowledge Sharing Point focusing on the intersection of Key Knowledge, Key Sharing Point, and Key Timing.

The third research gap identified in this paper is to find a better understanding of when exchange of key knowledge is necessary to support a process that both results in good design and consensus.

This paper has contributed to the understanding of how standards-making is situated in a multi-cultural, working group specific, and domain specific context. However, we would claim there is more research needed to fully understand how to design a process that will give an optimal result embarking upon a new work item in ICT standardisation for learning technologies.

6. Conclusions and Further Work

This paper is part of an ongoing reflection on our own practice in the field of design for learning technologies, in particular through international standardisation. We have established a background of what we would understand as best practices related to the management of standardisation processes that would support innovation, especially in creating anticipatory standards. On this background we have projected the case study of an ongoing project in the field of privacy and data protection for learning analytics. Based on this case study we have reflected on the relationship between the academic research
community, the standards community and the adopters of standards. The case study has contributed to identifying several gaps in our knowledge about the processes in question. These research gaps should be addressed in further research to be done as a continuous endeavour to improve design through action design research.

References


Abstract: This paper introduces a system that assesses how an elementary student acquires unfamiliar information to solve non-routine real life problems in the perspective of information processing. Students are required to understand the situations, select and apply the relevant information to solve the problems. All the problem solving procedures, including inquiry behaviors, are automatically recorded for off-line data analysis. By running an experiment with a total of 32 grade-5 students, our study evaluated their problem solving processes by analyzing the correlation between students’ performance and behavior data. The results reveal that students might have difficulty in applying the information to solve the problems correctly, and it suggests that integrating discrete information may be the biggest obstacle for the elementary students to solve the problems.

Keywords: problem solving, elementary students, information processing, behavior data

1. Introduction

How to solve non-routine problems is one of the most essential skills in 21st century (Griffin, McGaw, & Care 2012). Problem solving is the process of finding a means for reaching some goal from the initial state. Being different from well-educated adults, young students clearly have different levels of competences in solving complex domain-general problems (Findings 2014). It has fostered great demand for teaching this domain-general problem solving skill to them (Greiff et al. 2014). Fortunately, there has been research focusing on the assessment of problem-solving process. Dickison et al. (2016) assessed nursing clinical judgment by analyzing how problem solvers utilized symptoms of the simulated patient. Schweizer et al. (2013) introduced two instruments MicroDYN and MicroFIN which are proved to be the most well-established tools for assessing complex problem solving ability from the perspective of system thinking. To analyze students’ intention based upon behavior data, Evidence-Centered Design (ECD) has to be adopted (Mislevy, 1994), which defines the assessment framework to ensure the way in which evidences are gathered to be able to interpret the underlying assessment purpose. Based on all these previous researches, we aim to provide supports of problem solving for Chinese students who are usually better at mathematics, and worse at integrating discrete pieces of information (PISA, Publishing, 2010; OECD, 2016).

In this study, students are required to first understand the problem, then gather the relevant information to solve it. The key ability of problem solving in our case is to distinguish the information,
then apply the appropriate one. On the other hand, as our system does provide more than enough information, the type of information that a student accesses should reflect how the student value the different types of information (Johnson, Häubl, & Keinan 2007).

In this paper, we explained the results that show how elementary students solved problems by applying the relevant material. Our hypothesis is that higher percentage of relevant reading would always lead to better problem solving performance.

2. The Assessment System

The system provides a general framework for assessment tasks and implements the functionalities that support online student interaction recording. In the system, one assessment task could have many test items, and each test item can be either a multiple choice question, fill-in-blank question or an interactive question. Most of the test items cannot be solved without referring to their relevant information, but some standalone test items are existed correlating no relevant document.

To support the structure of the test items, the system provides some common functionalities and utilities. For example, there is a navigation bar placed to the right of the area where the test items display, like Figure 1. All the associated materials to the current assessment task are stored in a component called “material center”. Material center contains not only the relevant materials, but also some irrelevant ones. As soon as a student clicks on material center, a new window will popup, which shows the list of all the materials. When a student clicks on the name of a document, the detail will be displayed, like Figure 2. As soon as the student finishes a test item, he/she can proceed to the next one.

![Figure 1. The sample of a test item](image1)

![Figure 2. material center: The list of the materials (left); The content of a single material (right)](image2)

We conducted our experiment and data analysis with one representative assessment task. It uses camping as the story line and contains four test items, as Table 1. In the first test item, students are expected to read the relevant materials to decide which is the best way to purchase tickets. In the second test item, students are expected to correctly measure the length and width of three tents, then use the methods described in the relevant materials to calculate the capacity of the tents. To solve the third...
problem, students need to assign 54 people into 7 tents, without reading any relevant materials. In the last task, students are required to do a simple calculation for the consumption of water and food for the camping events referring to material center.

Table 1: Summary of test items

<table>
<thead>
<tr>
<th>Item name/Feature</th>
<th>Item type</th>
<th>Has relevant materials</th>
<th>Type of interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticket purchase</td>
<td>Multiple choice</td>
<td>Yes</td>
<td>Correct/Incorrect</td>
</tr>
<tr>
<td>Tent capacity calculation</td>
<td>Interactive simulation</td>
<td>Yes</td>
<td>Correct/Incorrect</td>
</tr>
<tr>
<td>Tent assignment</td>
<td>Interactive simulation</td>
<td>No</td>
<td>Progress/back-step</td>
</tr>
<tr>
<td>Path design</td>
<td>Fill-in-blank</td>
<td>Yes</td>
<td>Correct/Incorrect</td>
</tr>
</tbody>
</table>

The system is able to record every single interaction such as clicking on an alternative, and accessing the relevant materials and all the interactions are stored as JSON format into MongoDB. To standardize the interactions, all of them are labeled with 6 different types: correct, incorrect, progress, retreat, valid, and invalid. When the correctness of an interaction can be easily decided given the current context, the interaction will be labeled as correct or incorrect. Otherwise, an interaction will be labeled as progress or back-step based on whether the interaction move the state towards the goal. Some interactive simulations intentionally embed unnecessary interactions which do not help solving the problem at all. In this case, a necessary interaction will be labeled as valid, otherwise will be labeled as invalid.

3. Experiment Design and Analysis Models

The experiment is divided into two parts. First of all, students should go through the introduction task which helps students to become familiar with the User Interface. As soon as the introduction task was done, they started to work on the actual assessment task. Students needed to finish all the test items within 40 minutes independently. 32 elementary students in grade 5 took part in this experiment.

In order to understand how elementary students solve problems in the perspective of information processing and diagnose their issues during the solving procedures, we mined the relationship between information processing behaviors and students’ outcome performance in test items. In specific, we inspected their behaviors in terms of their interactions on material center. The behaviors were aggregated as the frequency of relevant and irrelevant reading, and we performed Pearson correlation to measure the relationship. We also conduct Hidden Markov Model (HMM) to infer students’ internal problem solving states by observing their behaviors.

The grading criteria is obvious for the first, the second, and the last test item. Students simply earn points when their answers were correct. However, the percentage of the interactions labeled as progress of a student is then used to represent his/her score in the third test item.

When a test item is an interactive simulation, it is possible to get a student’s in-progress correctness, which can be also considered as the correctness of a step (Vanlehn et al. 2007). Given this, we extracted this in-progress correctness for the second and the third test items.
4. Results

4.1 Descriptive Results

As mentioned before, there were 4 test items in total. Each item was weighted as 1 point. So the maximum possible score of the test was 4. On average, 36.7% students’ reading was relevant to solving the current assessment task. Their average score of the assessment tasks was 1.572 (SD=0.762). The average score and the standard deviation of each individual task was described in Table 2. The table also demonstrated the average frequency of students’ relevant reading and irrelevant reading behaviors to finish each item. The results indicated that the difference in how a test item should be answered didn’t seem to affect students’ performances. However, students performed much better when reading relevant information was not required.

Table 2: Descriptive results of the test items

<table>
<thead>
<tr>
<th>Task</th>
<th>Ticket purchase</th>
<th>Tent capacity calculation</th>
<th>Tent assignment</th>
<th>Path design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>0.5(SD=0.762)</td>
<td>0.25(SD=0.237)</td>
<td>0.68(SD=0.413)</td>
<td>0.136(SD=0.195)</td>
</tr>
<tr>
<td>Relevant reading</td>
<td>0.5(SD=0.622)</td>
<td>0.78(SD=0.608)</td>
<td>N/A</td>
<td>0.59(SD=0.911)</td>
</tr>
<tr>
<td>Irrelevant reading</td>
<td>0.91(SD=1.254)</td>
<td>0.63(SD=0.793)</td>
<td>0.15(SD=0.330)</td>
<td>1.09(SD=3.306)</td>
</tr>
</tbody>
</table>

4.2 Correlation Results

Pearson correlation coefficients were calculated to discover whether material reading behaviors can affect students’ performances. In specific, for each student, we first calculated the overall percentage of relevant material reading behaviors, and calculated Pearson correlation coefficient to check whether this percentage correlates the student’s overall performance and the performance in individual test items. The results were showed in Table 3. From the result, different reading behaviors didn’t lead to significantly different levels of performances. So we further explore whether their reading behaviors affect in-progress performance. It turned out that correctly referring to the relevant materials could lead to more valid interactions (r=0.403, p=0.022). Unfortunately, the valid interactions did not always end up with final correctness.

Table 3: Correlation between the percentage of relevant material reading and test item performance

<table>
<thead>
<tr>
<th></th>
<th>Ticket purchase</th>
<th>Tent capacity calculation</th>
<th>Tent assignment</th>
<th>Path design</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation</td>
<td>-0.009</td>
<td>0.185</td>
<td>0.305</td>
<td>0.089</td>
<td>0.156</td>
</tr>
<tr>
<td>significance</td>
<td>0.962</td>
<td>0.310</td>
<td>0.090</td>
<td>0.627</td>
<td>0.394</td>
</tr>
</tbody>
</table>

4.3 HMM Results

We set the number of hidden states as 3, which represented 3 different competence states. The emission probabilities of the corresponding observations, as listed in Table 4, were used to explain the meaning of the 3 hidden states. A student in H1 was probably bad in solving problems by reading additional materials, because students in this state often read irrelevant materials, and make invalid or incorrect
interactions. The observations of H2 were dominated by the behavior “Retreat”, and the observations of H3 were dominated by the behavior “Progress”. The 2 behaviors (“Retreat” and “Progress”) can be only observed in the test item “tent assignment”, which does not require students to read materials. A student in H2 should be bad at solving this planning problem. In contrast, a student in H3 should be good at solving this planning problem. Unfortunately, HMM model failed to detect a hidden state where students stay high competence for solving problems that require additional reading. It is probably because most of the students performed poorly during the assessment.

Table 4: The hidden states and their emission probabilities (drag stay)

<table>
<thead>
<tr>
<th>Observation /Hidden states</th>
<th>H1 (struggling in information processing tasks)</th>
<th>H2 (struggling in planning)</th>
<th>H3 (smooth in planning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>Invalid &amp; incorrect P=0.178</td>
<td>Retreat P=0.930</td>
<td>Progress P=0.959</td>
</tr>
<tr>
<td>O2</td>
<td>Incorrect P=0.163</td>
<td>Progress P=0.0696</td>
<td>Retreat P=0.0274</td>
</tr>
<tr>
<td>O3</td>
<td>Valid &amp; correct P=0.152</td>
<td>Invalid, incorrect P&lt;0.001</td>
<td>Long-time reading relevant material P=0.00161</td>
</tr>
<tr>
<td>O4</td>
<td>Invalid &amp; correct P=0.123</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The transition probabilities among the hidden states were illustrated in Figure 3. It means that the inferred competence states were relatively stable.

Figure 3. The transition probabilities among the hidden states

5. Discussion and Conclusion

The results showed that students did not perform very well in general. Furthermore, students exhibited low percentage of relevant material reading while performed much better in the task that did not require reading additional materials. It means that integrating discrete information may be the biggest obstacle for the students to solve the problems. The results from the reading section in PISA 2009 (Publishing, 2010) also reported similar findings.

Both of descriptive and HMM results imply that processing outside information to solve problems stands as a relatively independent skill. From the results of HMM analysis showed the
students who were good at solving planning problems did not tend to make either relevant or irrelevant readings. In addition, HMM analysis was only able to figure out the hidden state from the data that mainly lead to irrelevant reading and incorrect behaviors. The results again implied that students had difficulty in acquiring and applying relevant information for solving problems. But students did not completely fail in solving the problems. The descriptive result showed that students did go to check the reading materials frequently while necessary, and seldom went to material center when the task did not require, however, at most of the time, students were not able to distinguish which material was relevant. Therefore, students might be able to recognize when they need to refer to additional material, but were at low competence of acquiring relevant information for solving problems.

Acknowledgements

This work was supported by Chinese Ministry of Education Humanities and Social Science Foundation under projects number 14YJC880025, and China Postdoctoral Science Foundation under projects number 2017M610054.

References


Algorithm Learning by Comparing Visualized Behavior of Programs

Daiki IHARAa, Satoru KOGUREb, Yasuhiro NOGUCHIC, Koichi YAMASHITAD, Tatsuhiro KONISHIb & Yukihiro ITOHd

aGraduate School of Informatics, Shizuoka University, Japan
bFaculty of Informatics, Shizuoka University, Japan
cFaculty of Business Administration, Tokoha University, Japan
dShizuoka University, Japan
*ihara.daiki.17@shizuoka.ac.jp

Abstract: It is said that visualizing the behavior of programs is an effective way for novice programmers to understand algorithms. Based on discussions with teachers of programming subjects, it was suggested that learners could deepen their understanding of algorithms by comparing visualizations of programs. This paper proposes learning processes to understand algorithms by comparing two figures. In addition, we develop a system that provides supporting functions for comparisons. Our experimental results show that the learning process using our systems is an effective way for novice programmers to understand algorithms.

Keywords: Program visualization, behavior comparison, algorithm understanding.

1. Introduction

Developing skills for understanding the behavior of programs is important in programming education. It is said that visualizing the behavior of programs is an effective way for novice programmers to understand algorithms. It is better to demonstrate the mechanism of the given algorithm in words and pictures than only in words (Gabor, 2009). Therefore, many program visualization tools have been developed (Rößling et al., 2002, Yamashita et al., 2016, Moreno et al., 2004, Fossati et al., 2008, Malmi et al., 2004). Rößling developed ANIMAL, which shows the behavior of programs based on the ANIMAL SCRIPT written by teachers. Teachers can achieve visualization at low cost by using preset sample scripts. Yamashita developed TEDViT (Teacher Explaining Design Visualization Tool), which has more distinctive features than existing program visualization tools. By using TEDViT, teachers can control the visualization of the unique concepts of algorithms (e.g., tree node and stack) and the behavior of programs. Teachers can present program visualization to their students based on their teaching design. If needed, teachers can use the figures from their classroom textbooks with TEDViT. The figures are called the Status of Target Worlds (STWs). To use TEDViT, teachers must write rule sets. The rule sets are then referred to as drawing rules. By using the tool, learners can trace the steps of programs with program visualizations.

Through discussion with teachers of programming, it was suggested that learners could deepen their understanding of algorithms by comparing the STWs of two scenes (scene means view of specific point in time). Through the discussion, we thought that learning by comparing scenes could be classified into three viewpoints: (A) Compare different time points, (B) Compare different types of data to be processed, and (C) Compare different programs. These details are described in section 2. Learning by comparing scenes is referred to as Scenes Comparison Based Code Reading (SCBCR).

Above mentioned program visualization tools do not support such comparison of multiple scenes. Therefore, the aim of this work is to construct a system that supports SCBCR. We extend the current TEDViT to support SCBCR. The target audience of the system includes learners who do not attain an adequate understanding after a lecture class. It is difficult for such learners to find scenes that should be compared. Therefore, a teacher should be able to designate scenes suitable for SCBCR and to make messages that advise learners what they should focus on in SCBCR. We extend the existing drawing rules to enable teachers to realize SCBCR.
2. Fundamental Consideration

2.1 Classification of Comparison

Based on discussions with teachers of programming subjects, we classified SCBCR into three viewpoints based on comparison targets. In this section, we describe SCBCR in detail and discuss the three viewpoints based on the comparison programs defined by teachers.

(A) Comparing Different Time Points (Viewpoint (A)): Learners compare scenes that correspond to two different points in time during the execution process. The learners understand the behavior of programs from the two scenes at different time points. As an example, in the process of a sort program, the learners can understand the step of sorting by comparing the scene of the array up to the nth element with the scene of the array up to the (n+1)th element.

(B) Comparing Different Types of Data to be Processed (Viewpoint (B)): Learners compare two scenes in the same program with different data. Through this comparison, learners understand the program operation independent of data, and the difference in execution efficiency due to the difference in data. As an example, in a sort program, the learners compare the execution process of sorted data with data in reverse order. This allows them to observe the difference in the sorting process with different data.

(C) Comparing Different Programs (Viewpoint (C)): Learners compare scenes from the execution of two different programs. This type can be further classified into the following three categories:

- **Programs with the same purpose but different algorithms**: Learners can understand the difference in execution efficiency due to different algorithms.

- **Programs with the same algorithm but different implementations**: Learners can understand the difference in execution efficiency due to different implementations.

- **Correct program, and program that have bugs**: Learners focus on the behavior of the buggy programs compared with the behavior of the correct programs. This comparison leads learners to a better understanding of the behavior of the correct programs.

2.2 Required Functions for Our Proposed System

The existing TEDViT supports the visualization of only one program. To realize SCBCR, TEDViT requires a new interface design and configurable rules for comparing two different scenes. In this paper, these configurable rules are referred to as **rules for supporting SCBCR**.

In SCBCR, the learners compare two scenes, so the system needs to show two scenes. In SCBCR with viewpoint A, there is one execution process. In SCBCR with viewpoints (B) and (C), there are two execution processes. We develop an SCBCR system for viewpoint (A) and an SCBCR system for viewpoints (B) and (C). In SCBCR, learners sometimes compare two scenes in single execution process. The interface supporting SCBCR should show the two scenes in sync to the step of the program. The learners can easily step forward and backward through the scenes. In this paper, the function is referred to as **the step observation function**.

Teachers should be able to configure rules for supporting SCBCR for the comparison of two scenes. To realize an effective comparison, two types of rules are needed. First, teachers set the scenes to be compared by learners. In this paper, the first rule type is referred to as **saving scenes rules**. Second, teachers set up comparison viewpoints with two scenes and add an advisory message. In this paper, the second rule type is referred to as **design comparison rules**. To design these rules, we analyzed some algorithms that were used in programming lectures.

3. Design of Systems and Rules for Supporting SCBCR

3.1 Proposed System for SCBCR with Viewpoint (A)

Figure 1 is the system overview for SCBCR with viewpoint (A). It consists of five areas.

**Program code area (Area 1)**: The system shows the program code in this area. The system emphasizes a statement whose scenes are shown in the system. The system supports two scenes identified by red
and blue. Two scenes corresponding with a statement are highlighted by color gradation from red to blue.

**Buttons area (Area 2):** The system creates buttons for operating the system. Learners proceed with SCBCR by using these buttons. The system normally shows two scenes defined by rules for supporting SCBCR. If learners would like to check the steps between the two scenes, they use the change mode button. The system then shows the first scene and the scene in the next step.

**Scene area (emphasized by red or blue) (Area 3 and 4):** Area 3 and 4 shows the scene corresponding with the statement highlighted in red or blue in area 1.

**Message area (Area 5):** The system shows messages defined by design comparison rules.

The scenes in Figure 1 are ones comparing the nth with the (n+1)th laps of a for-loop that searches for a character string in a longer character string by using the KMP algorithm. Learners compare two scenes in the execution process to grasp how the state changes in one lap of the for-loop.

![Figure 1](image1.png)

**Figure 1.** An overview of the system for viewpoint (A).

### 3.2 Proposed System for SCBCR with Viewpoints (B) and (C)

Figure 2 is the system overview for SCBCR with viewpoints (B) and (C). It consists of eight areas.

![Figure 2](image2.png)

**Figure 2.** An overview of the system for viewpoints (B) and (C).

In the SCBCR with viewpoint (B) and (C), there are two execution processes. In this system, one process is shown in area 1, area 3, and area 5 on the left side. Another process is shown in area 2, area 4, and area 6 on the right side.
3.3 Proposed Rules for Supporting SCBCR

To save scenes which should be compared by learners, teachers write saving scenes rules. In the rules, teachers can tag the scene with a unique label name. To designate scenes for comparison and messages to advise what should be focused on in SCBCR, teachers write design comparison rules. In these rules, teachers specify tags for the scenes to be compared, and compose the messages. We have to omit these description methods for want of space.

4. Experimental Evaluation

We conducted an experiment to confirm the effectiveness of our two support systems. The hypotheses of the experiment are that our system for SCBCR is useful for learners as following:

- Learners can cultivate better understanding by SCBCR supported by our system.
- Learner’s understanding of processes is related to comparison scenes.
- Learners perform a long and thoughtful observation of comparison scenes.
- There is no significant problem in using SCBCR with our system’s interface and the messages.

4.1 Methods of the Experiment

We recruited experiment participants who had C language knowledge and had attended programming lectures. Seven university students participated in the experiment. Six of them were in the faculty of informatics and one was in the faculty of engineering. They had a half-year or more of experience attending C language programming lectures. We employed four algorithms for learning: the KMP algorithm, the insertion sort, the bubble sort, and the selection sort. The informatics participants had knowledge of all four algorithms, and the participant of engineering had knowledge only of bubble sort. In this experiment, the participants proceeded with three types of SCBCR.

**SCBCR 1 (the KMP algorithm with viewpoint (A)).** This includes two points we want learners to understand. The first is that to process the contents of a for-loop is to proceed with string search by changing the position of interest in the search target strings and pattern character strings. The second is that the focusing position of the pattern character string may change according to the table.

**SCBCR 2 (the insertion sort algorithm with viewpoint (B)).** This includes the content that we want learners to understand; that is, in the case of reverse order data, the calculation time required for sorting is increased compared with sorted data.

**SCBCR 3 (the bubble sort and the selection sort algorithms with viewpoint (C)).** This includes the content we want learners to understand is that although the number of comparisons between selection sort and bubble sort is the same, the number of exchanges in selection sort is less than in bubble sort. Therefore, selection sort is more efficient.
We conducted the experiment using the following procedure.

1. Answer pre-test questions for SCBCR1, 2 and 3.
2. Receive an explanation on how to use the system for SCBCR with viewpoint (A).
3. Receive an explanation on SCBCR 1 with viewpoint (A) and proceed with that.
4. Answer post-test questions for SCBCR 1 with viewpoint (A) and operation test for SCBCR 1.
5. Receive an explanation for how to use the system for SCBCR with viewpoints (B) and (C).
6. Receive an explanation for SCBCR 2 with viewpoint (B) and proceed with that.
8. Receive an explanation for SCBCR 3 with viewpoint (C) and proceed with that.
10. Answer the systems questionnaire for SCBCR and messages for comparison.

The pre-test in (1) consists of three questions; the first one is about the KMP algorithms, the second one is about the insertion sort, and the third one is about the bubble sort and the selection sort. The post-test consists of the same three questions as the pre-test. The explanation for SCBCR 1, 2, and 3 does not contain the answers for the pre-test and post-test. In (4), we give participants an operation test. In the operation test, the participant answers to supervisor with pairs of scenes that perform specific processes indicated by the question. The questionnaire in (10) consists of seven questions about the evaluation of the systems interface and comparison messages. The questionnaire items are evaluated on a five-point grading scale, with a higher number indicating a better rating.

After experiment, we compare answers for the pre-test and the post-test. We analyze whether there is an improvement in the answers about SCBCR, including what we want learners to understand. In the case that there is such an improvement, we conclude that SCBCR has shown learning effects. If the participant answers with the correct pairs of scenes in the operation test, we can evaluate the learner’s understanding related to the comparison scenes shown by the system.

4.2 Results of the Experiment

4.2.1 Analysis of Test Results

We compare the answers of the pre-test and post-test. Table 1 shows scores of the three questions. The values in the column are the number of informatics students, engineering student.

For the KMP algorithm, we give participants an operation test. The operation test has three correct answers. All of informatics participants gave two or more correct answers (Three participants obtained full marks). In addition, the engineering participant also obtained full marks. The results suggest that the learner’s understanding of process is related to the comparison scenes shown by the system. The facts on Table 1 and operation test suggest that SCBCR with viewpoint (A) has some learning effect. Also the facts on Table 1 suggest that SCBCR with viewpoint (B) and (C) have some learning effects.

From the above results, it can be concluded that SCBCR with our systems has some learning effects. In addition, since SCBCR is also effective for a participant of engineering, it suggests that SCBCR with our systems is effective for various knowledge levels of learners.

Table 1: Scores of the three questions (Informatics students, Engineering student).

<table>
<thead>
<tr>
<th></th>
<th>KMP algorithm</th>
<th>Insertion sort</th>
<th>Selection sort/bubble sort</th>
</tr>
</thead>
<tbody>
<tr>
<td>correctly answered in the pre-test</td>
<td>2, 0</td>
<td>4, 0</td>
<td>0, 0</td>
</tr>
<tr>
<td>improved their answer in the post-test</td>
<td>4, 1</td>
<td>2, 1</td>
<td>6, 1</td>
</tr>
<tr>
<td>did not improve their answer</td>
<td>0, 0</td>
<td>0, 0</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

4.2.2 Analysis of Operation Logs

The system for SCBCR collects operation logs of learners. We define irregular comparison behavior in SCBCR as a learner proceeding with successive comparisons in a short time in comparison scenes that the learner has not observed yet. We searched for irregular comparison behavior in the operation log of the learners. In SCBCR with viewpoint (A) and (C), no one showed irregular behavior. In SCBCR with
viewpoint (B), three informatics students forwarded comparisons continuously, with less than one second between comparisons in certain circumstances. We guess they do not need to observe for a long time because they understand the target algorithm well. These facts suggest that the system almost be used as we intended.

4.2.3 Analysis of Questionnaire Answers

Table 2 shows the results of the questionnaire to evaluate educational effectiveness of the system. As for (A), the results suggest that the participants evaluated the system highly. As for (B) and (C), the first question suggests that the participants evaluate highly, but the evaluation with viewpoints (B) and (C) is lower than (A) in the second and third questions. For messages from the system, the participants’ evaluation was high with an average of 4.00. One participant commented that it was easy to proceed with SCBCR because messages of comparison were brief and easy to understand. These results suggest that there is no significant problem in our system’s interface and messages.

From the results in 4.2, we conclude that SCBCR is effective for learners to understand algorithms and that our systems can support learners for SCBCR.

Table 2: Results of the questionnaire.

<table>
<thead>
<tr>
<th>Question</th>
<th>(A)</th>
<th>(B), (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How easy did you feel to understand the content of the component elements on the interface?</td>
<td>4.14</td>
<td>4.14</td>
</tr>
<tr>
<td>How intuitive did you feel the arrangement of component elements on the interface was?</td>
<td>4.14</td>
<td>3.43</td>
</tr>
<tr>
<td>How easy did you feel the system’s operation?</td>
<td>4.43</td>
<td>3.71</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, we proposed comparison learning processes called SCBCR. We developed systems for SCBCR by extending TEDViT. The results of our experiments suggest that our systems for SCBCR are effective in helping learners understand algorithms. Going forward, we will improve the system based on the results of a questionnaire and the opinions of teachers.

Acknowledgements

This research was supported by the Japanese Grant-in-Aid for Scientific Research (B) Grant Number JP24300282, (C) Grant Number JP16K01084 and for Young Scientists (B) Grant Number JP15K16104.

References


Analyzing the E-learning Video Environment Requirements of Generation Z Students using Echo360 Platform

Swapna GOTTIPATI* & Venky SHANKARARAMAN
School of Information Systems, Singapore Management University
*swapnag, venks@smu.edu.sg

Abstract: As with any other generational cohort, Generation Z students have their own unique characteristics that influence their approach to learning process. They are the future workforce and several efforts are undertaken by Government and education institutes to consider the characteristics of Gen-Z in developing the a curriculum and teaching environment suitable for these students. E-learning plays a key role in students learning process and has been widely adopted by many education institutions. In particular, videos play a major role in the learning process of Gen-Z students. The purpose of this paper is to focus the on requirements of Gen-Z students and to provide suggestions for how to create a e-learning video environment for them. This study aids the education institutes in designing and developing e-learning environment appropriate for Gen-Z students. For this project, we study the e-learning videos environment, Echo 360, and the behavior of Gen-Z students to suggest the additional features to improve the e-learning video environment. Around 76 students participated in our study where they had to complete an e-learning session and a qualitative survey to provide feedback about the session. Our study shows that the students appreciate the video based learning due the flexibility, pace and independence. At the same time, they prefer online interactions, and provided technology based suggestions for e-learning environment improvements.

Keywords: e-learning video environments, Echo 360, Generation Z students, requirements

1. Introduction

E-Learning has become a key learning methodology in education systems. The online tools should be designed to enhance the flexibility and accessibility of the teaching and learning experience of e-learning users. Critical to successful and effective e-learning is understanding the users of the tools and their requirements (Shili et al. 2016; Ivanova & Smrikarov, 2009). Generation Z (Gen-Z) students have been raised in a world of instant access to knowledge and information, a world of automation, remote controls, and simulation capabilities to stimulate the mind. They prefer non-traditional teaching methods and are far more technologically savvy. The students believe in smart work and not hard work. Therefore, the institutions should cater for the requirements of the Gen-Z students for the success of e-learning education methods.

1.1 E-Learning

E-learning is a methodology widely practiced in education institutes to deliver the teaching and learning experiences through technology (Shili, 2016).

Advantages of e-learning include convenience, cost, higher retention, greater collaboration, global opportunities, flexibility and mobility. Some disadvantages can be lack of control, isolated, learning approach and technology issues (Olliges & Mahfood, 2003). Nevertheless, e-learning is highly popular and adopted by several learning institutes around the world.

The majority of students who enter higher education are required to use online learning resources or activities to support their formal or informal learning. Examples include virtual learning environments, discussion forums/lists, e-mails, podcasts, or library information databases. Content
types widely used in e-learning can be categorized into text (Wikis), documents (slides, pdf, docs, xls) and videos (Kinshuk & Nian-Shing, 2006; Stefan, 2007).

Central to e-learning success are communication technologies, content frameworks and methodologies which are generally categorized as synchronous or asynchronous. Asynchronous e-learning is a flexible online learning methodology that is often facilitated by tools such as e-mail and discussion boards. It supports learning process and network among learners and with teachers outside the constraints of time (Kinshuk & Nian-Shing, 2006; Hampton & Keys, 2017). Synchronous e-learning requires the learners and teachers to be online at the same time. It is commonly supported by media such as videoconferencing and chat. Learners experience synchronous e-learning as a real class room environment with the key feature of asking and answering questions in real time (Stefan, 2007)

Successful e-learning experience will use a combination of technologies that can integrate various components of the learning process namely; instructors, students, course content and assessments. Within the higher education and e-learning fields there is a growing level of interest in exploring and understanding the e-learning tools and experiences of students in universities (Olliges & Mahfood, 2003; Zhang & Nunamaker, 2003).

1.2 Millennials and Z Generations

Millennials or Gen-Ys are the people born between 1978 and 1994. Most of the undergraduates, graduates and significant number of faculty belong to this cohort. Most of them are considered as first generation technology users and they experienced world with computers and internet since childhood. Their world is different from previous generations in the aspects of exposure to technology (Michael & Richard, 2004; Reynol & Jeanna, 2006). Considering these differences, previous works offered specific suggestions for meeting the needs of millennial and Gen Y students; provide cutting-edge technology, interactive web services and an infrastructure for virtual communities; and offer additional technological support for these first-generation technology driven students (Tracy, 2007; Reynol & Jeanna, 2007). Learning shifted from instructor-centered to learner-centered, and is undertaken anywhere, from classrooms to homes and offices. With the availability of advanced information and communication technologies, e-learning had a far-reaching impact on learning millennials (Zhang & Nunamaker, 2003; Jones et. al, 2007; Ivanova & Smrikarov;2009).

Generation Z students are cohort after millennials or those who were born in 1995 or later. They are sometimes referred to as ‘post-millennials’ and they have started appearing in Universities (Hampton et al., 2016). There are many key differences between millennials and Gen-Z and therefore there is a burning need for redesigning education practices to suit the Gen-Z characteristics (Loehr, 2015).

Gen-Z students view education as a necessity for individual success and societal prosperity. Although Gen-Z students share some of the same characteristics of millennials, they have their own unique characteristics (Loehr & Generation 2015). The cohort can be described as compassionate, thoughtful, determined, and responsible (See Miller et al., 2016). A significant aspect of Gen-Z is the widespread usage of the Internet from a young age. Members of Gen-Z grew up with technology and addicted to digital devices. They spend more time online interacting on social media websites for a significant portion of their socializing. These students prefer non-traditional teaching methods, and like using logic-based approaches and experiential learning. (Kinshuk & Nian-Shing, 2006) “How to” videos, online encyclopedias, and multiple other resources are available for students to access to help them learning (Hampton & Keys, 2017). According to the recent study of Microsoft Corp, new generation have an 8-second attention span almost close to the attention span of a gold fish. At the same time, 11 percent of them are diagnosed with attention deficiency syndrome (MCS, 2017). Other studies also show that Gen-Z are less focused and prefer individuality (Hampton & Keys, 2017). Therefore, it is crucial to study and analyse the requirements of the Gen-Z students in designing the e-learning tools else they will lose their interest, leading to the failure of e-learning efforts.

In this paper, we study the perspectives of Gen-Z students on the e-learning video environments. We want to capture new ideas from students that can benefit them more in learning process. Most importantly we are looking at e-learning video environment for Information Systems Gen-Z students. Our research approach is based on Echo 360, a learning video platform for education. The outcome of this study aids instructors and instructional designers to develop more efficient and effective instructional methods based on student learning style. Furthermore, the project outcomes aid in improving the e-learning features to fit various learning approaches of Gen-Z students.
The rest of the paper is organized as follows. In section 2, we present the research methodology used in our project. Section 3 describes the experiment setup for data preparation, and quantitative and qualitative analysis of our research results. Conclusions drawn from this research are presented in Section 4.

2. Research Methodology

This research study is conducted on undergraduate students from School of Information Systems, Singapore Management University. We studied their e-learning requirements using third year course, Enterprise Web Solutions. The course is conducted for 15 weeks in the normal class room settings. We introduced e-learning session for one of the weeks. The methodology involved three main stages.

In the first stage, the course instructors prepared videos for topic, “Analytics in Enterprise Web Solutions”, using Echo 360 (Fei et al, 2017). Echo is a tool that enables to create videos and at the same time provides the statistics of the student participations. Echo360 lecture capture system provides recording choices based on the curriculum, instructor preference and venue. The tool provides features to create videos, editing videos and publishing videos online. It provides students with an easy option to access course materials online from anywhere and at any time. It has features such as; pause, play, share, discussions, bookmarks, feedback etc.

The instructors of the course prepared short videos, 15min to 20mins. Short videos are useful for Gen Z students who are less focused and low attention span (Hampton & Keys, 2017). Each video contains lecture content followed by exercises to apply the learning. This arrangement enables to address their individuality and challenging nature.

In the second stage, we posted the videos to e-learn, SMU’s learning management system. The students are required to complete the e-learning within the week.

In the third stage we surveyed the students using online survey form and collected the results. We also collected the statistics from Echo on various aspects such as students’ video views, completion of view for each student and date. The survey was designed to identify the following:

1. What are the likes and dislikes about the e-learning and videos?

2. What are the suggestions they would give for the improvements of e-learning session and videos?

Likert scale surveys are useful to know the numerical scale of sentiments. However, the opinions and suggestions on the videos which are expressed as plain text are critical to understand the requirements of the students. Our objective is to analyze Gen-Z students’ thoughts and apply for the e-learn improvements. Therefore, we created the survey as open ended questionnaire to capture textual comments. Finally, we analyzed the qualitative results from the survey and quantitative statistics from Echo to understand the behavior and the e-learning video environment needs of the Gen-Z students.

3. Experiments and Findings

In this section, we first present data preparation followed by the details of results. We present our quantitative analysis based on the statistics from Echo. We then present the qualitative analysis of the survey data. We finally present the summary of the e-learning video environment requirements of Gen-Z students.

3.1 Data Preparation

Over 76 students participated in the study, and we collected Echo data and survey data to analyse the e-learning needs for Gen-Z students. The students are of age group between 18 and 21, who are born in and after 1995. 76 students participated in e-learning during week 6 of the course delivery. The students completed e-learning session out of class and then participated in the survey.
3.2 Background of Videos

The course topic is on analytics in enterprise web. Three videos are created to give an overview of analytics, its role in enterprise web and the tools useful for embedding analytics in the enterprise web. After each video, the students will be given an online quiz to test their understanding. The content of the video includes both theory and application-based learning components, which is a general pedagogy in courses under Information Systems program.

3.3 Echo Statistics

Recall that the instructors created three videos and published to SMU’s online Learning Management System, e-learn. We collected the statistics about the individual students’ data such as; amount of participation, time and date of participation, completion status of the videos, bookmarks, unique views and cumulative (total number of views) views and download status. The tool also enables to collect the overall course statistics by weekly.

Figure 1 shows the overall views. The views on November 11th represent the test video created by the instructors for training purpose before the semester. The instructors created three videos related to the course content during January. The actual course started in January and the views in February depict students’ participation in e-learning week. For all three videos, the total number of unique views is 127 and total number of cumulative views is 154. We observe that total number of students’ cumulative views is more than total number of unique views. Note that students do not need to completely view the video to capture these statistics. The cumulative views are more than unique views for several reasons.

Firstly, the students might have viewed the videos partially and revisited to view the remaining part. Secondly, the students might have viewed the videos multiple times for better understanding. We observe that the views on video 1 are higher than the other two videos – video 2 and video 3. This also shows that the Gen-Z students are uninterested in viewing the other videos. Therefore, it is necessary to understand their issues with the videos and the environment. Survey questions and analysis will address these issues in sub section C. To further understand the behaviour of the students, we analysed the views by week. Figure 2 shows the weekly statistics of the video views.

We performed a deeper analysis on the statistics and observed that eight students watched the videos for the second time during March and April. In our analysis, we observed that the final exam is in the first week of April. This shows that e-learning videos are helpful for Gen-Z students to revise for the final exam. To study the participation of the students, echo provides duration statistics. Table I shows the participation of the students on all three videos.

Echo statistics from Table 1 shows that the Gen-Z students used e-learning videos for learning and revising for exams. However, many of them they didn’t view all the videos or in other words, they didn’t complete the e-learning. Many students have ended the e-learning session after watching the first
video. Average completion of video 1 is 77.3% which is higher than video 3, 50.1%. To study the reasons for this behaviour, we designed the survey as explained in Section 2.

Table 1: E-learning video completion statistics

<table>
<thead>
<tr>
<th>Video Title</th>
<th>Average Completion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS305 - Week 6 - Video 1</td>
<td>77.32</td>
</tr>
<tr>
<td>IS305 - Week 6 - Video 2</td>
<td>60.12</td>
</tr>
<tr>
<td>IS305 - Week 6 - Video 3</td>
<td>50.14</td>
</tr>
</tbody>
</table>

3.4 Survey Statistics

In echo statistics analysis, we observe that the students didn’t complete the e-learning session and most of them ended the session after video 1. We observed that the survey questions based on Likert scale are not effective when collecting the feedback from students. Moreover students tend to provide scoring randomly. At the same time, we wanted them to provide us an effective feedback by using a simple survey form with qualitative questions only. We asked three questions in the online survey form to students to understand their thoughts on e-learning video environment, Echo360.

1. What you like about e-learning videos?
2. What is that you didn’t like about e-learning videos?
3. How can we improve? Provide suggestions.

The responses are textual in nature and hence we use word clouds to analyse and summarise the responses. Figure 3 shows the responses of likes and Figure 4 shows the responses for dislikes.

We observe that the students like about e-learning factors such as independence, time flexibility, pace, repetition, replay, different learning speeds, availability, convenience, and venue flexibility. These are the common aspects of Gen-Z students as studied by previous researchers. To study the issues on why they didn’t complete the e-learning sessions, Figure 4 aids us to analyse their dislikes and needs. We note that though they are independent learners who appreciate online learning, they also look for interactions and availability of professors to answer their questions and clarify their doubts. Table 2 shows the suggestions for improvements to e-learning tool.

Table 2: Suggestions for improvements in e-learning videos

<table>
<thead>
<tr>
<th>Category</th>
<th>Student Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content and delivery improvements</td>
<td>1. Post Q&amp;A sessions, 2. Instructor on call, online, 3. More exercises, quizzes and labs, 4. Faster pace, 5. Videos with professor teaching and not just slides, 6. Flip classroom, 7. e learning before lessons and labs, and discussions during lessons</td>
</tr>
</tbody>
</table>
The analysis of the third survey question, “suggestions”, is depicted in Table 2. We categorised these suggestions into tool improvements, and content and delivery improvements. We observe that Gen-Z students provide suggestions which are technology based. They provided innovative ideas based on their day to day life experiences. Suggestions such as discussion forums, chats, flip-class rooms show the tech savviness of the Gen-Z students.

4. Conclusion

Gen-Z students appreciate online learning education methods and the same time, they also expect more sophisticated tools that can provide online interactions, reminders, concurrent chat sessions, notes taking features and better site navigation designs. At the same time, they look for faster pace and short videos to balance their focus levels and lower attention spans. Our study summarizes the e-learning video environment requirements for Gen-Z students. The research provides suggestions on additional features to e-learning video environments to align with the Gen-Z student needs and the success of the e-learning sessions.

References

Tracy Knofla, (2007) “Understanding Today’s College Student and Providing Excellent Service” (presented at California State University, Long Beach, October 2007).

396
Train-For-Life: On-Line Interactive Training for Industry Learners
Bashar BARMADA & Nilufar BAGHAEI
Department of Computer Science, Unitec Institute of Technology, Auckland, New Zealand
bbarmada@unitec.ac.nz

Abstract: Learning a topic online via Massive Open Online Courses (MOOCs) has gained a lot of popularity due to their scalability and efficiency of knowledge distribution. However, participating in these courses usually means fully committing to the course for at least a few months in some cases. Providing engaging online education is a challenge for industry/non-academic trainees mainly due to their lack of time. In this paper, we describe a practical experience of online learning for non-academic audience in transport, logistics, security and safety industry. The proposed training courses are designed to be short, flexible and interactive to keep the trainees interested and engaged with the content. During any courses the trainees were challenged with different interactive tasks to emphasis on some important points. Results show that more than 90% of the trainees managed to complete the training parts of the courses they are registered to do and enjoyed their experience.

Keywords: online learning, interactive training, non-academic trainees, industry learners

1. Introduction

Training and development is a continuous process in any organization, in order to keep the employees on the top of their working field. With the permanent availability of the Internet, on-line training becomes very popular in offering cost effective learning materials anywhere and anytime [Kim et al., 2014; McCutcheon et al., 2015; Seaton et al., 2014; Gaebel, n.d.; Cisco Networking Academy, n.d.]. It is essential for companies to train different types of staff in transport and logistic industry. The aim is to increase their awareness of how to deal with difficult and emergency situations that will affect their health, safety and the security of the products and organisations, such as Lorry Hijacking, Deception and Rubber [Burke et al., 2011; Robson et al., 2012; BBC news, ret. 2012; New Zealand Police, ret. 2017].

Academic on-line courses can run for long hours, days and even weeks and usually the learners are enthusiastic to finish the course undertaking to improve their learning curve quickly and efficiently. However, providing engaging online education is a challenge for industry/non-academic trainees mainly due to their lack of time [Aamodt G. M., 2016; Robson et al., 2012; Kizilcec et al., 2013]. Our own experience also shows that some of these learners do not find the lengthy online courses engaging enough and see them as an obstacle to do the real work.

We propose Train-For-Life (T4L), an on-line interactive platform training courses targeting industry trainees in transport and logistic area (http://www.trainforlife.co.uk). The courses also target admin staff and their daily life, at work or outside work, to make it healthier physically and emotionally (HSE – Health and Safety Execution, www.hse.gov.uk/index.html). With on-line availability of these courses, drivers, admin staff and others are flexible to take the required training anytime and from any place. All courses are built with the following goals in mind:

- Courses are short (30-50 minutes long), as we are targeting industrial participants, who most of the time find it very difficult to find large gap in their schedule to do training, and who get bored and/or distracted very quickly.
- Courses are interactive. Beside simplicity, the courses are built with many interactive features to engage the participant and improve their perceptions (Harris, n.d.).
- All courses are followed by short tests of multiple choice questions to test the understanding of participants.
The research question we are investigating is given the online nature of these short courses and the audience (non-academic industry trainees), will they be able to complete these courses?

This paper is structured as follows. Section 2 presents the T4L platform and gives a brief description of the available training courses, the areas they cover and the mechanism to follow to create new courses. Section 3 describes the interactive methods used in the training courses. The results are presented in Section 4. We conclude by detailing the future work.

2. Training Courses

2.1 Courses Categories

Train-for-life develops the online training courses according to the following main categories:

- Transport and Logistic: Concentrates on drivers and how to improve their security awareness, for example in case of hijacking and theft.
- Personal Security: Focuses on how to improve the person’s own security at work and during everyday life routine, such as scam awareness.
- Security and Protection: Concerns on the safety of individual and the precautions that are needed to be taken, such as manual handling of heavy objects, and robbery response.
- Security Skills: Enhances the skills related to security and various awareness issues, such as conflict and stress management.

2.2 Tests

After finishing the learning part of any course and to ensure that the trainee has absorbed the essential concepts, the trainee needs to take a short test of 20 questions maximum with multiple choice answers to complete the course. To maintain the fairness of the test, the questions are chosen randomly from a large bank of questions. The results are recorded in the management system together with other important information such as the date of the test and start/end time of the test. After finishing the test, the trainee can revise the questions that been answered incorrectly, but they cannot change their results. The trainee can take the test of any course as many times as they want, independent of the learning material, until they reach the passing results.

2.3 Course Creation Mechanism

![Figure 1. The stages of developing the online training courses.](image)

The creation of the courses goes through several stages as illustrated in Figure 1. The stages are summarized as follows:

2.3.1 Requirement Analysis

Define a clear objective of the course where it gives benefit to the industry and provides training to a large sector of audience, especially if the course targets government concerns and is supported by professional bodies, such as health and safety executive. More thorough
research is conducted around the topic and many stakeholders are consulted at this stage to ensure the analysis satisfies the rules and regulation and the content delivers the right information. The interactive parts are also planned to enrich the course and make it more interesting to the type of audience.

2.3.2 Design

The course is divided into several short sections, each section consists of few pages followed by the interactive parts. The content of each page is a mixture of text, images and video, where the text does not occupy more than 33% of the page. Figure 2 shows an example of a used page template. Paragraphs appear sequentially and are synchronized with the images/videos. The reader has the option to hover over any text to view the related image or video. A progress bar indicates to the reader how far they are from finishing the contents of the page.

![Figure 2. Example of page template. Example: Conflict Management.](image)

Designing the interactive parts happens at this stage. It involves shooting video clips and taking photos, which are done using professional models, actors and photographer. All video clips and photos are edited then to suit the course under development.

2.3.3 Implementation

Where everything comes together. The course is implemented using PHP, Javascript, and JQuery for the interactive parts, which will be explained below.

2.3.4 Testing and Validation

This includes technical and non-technical verification and validation, and involves customers and different stakeholders for the final acceptance phase.

2.3.5 Maintenance

To ensure that the course is keep fulfilling the requirements of industry and professional bodies, the course is subject to continuous update, especially when feedback starts to come from users.

3. T4L Interactive Methods and Architecture

The interactive features used are what distinguish this online training and make the learning method unique and interesting, compared to other conventional learning methods. After each learning section the trainee will go through some interactive parts to challenge their understanding and reiterate on some
important points in the learning material for that section. The trainee can try the right way as well as the wrong way, and with every choice they make a brief explanation appears to comment on their action and the consequences that might happen. It is always advisable that the trainee try all the wrong ways to see the consequences. The used interactive methods are listed below:

- **Panoramic training:** The trainee can take a tour inside a building or a room, move between buildings/rooms, and click on points of interest to emphasis on some important ideas.
- **Drag and Drop:** The trainee can drag some objects and drop them in their right or wrong places, or sort objects in a particular order.
- **Using arrows:** The trainee clicks on different arrows (up/down, front/back, and right/left) to change the value of some parameters and experience the right and wrong ways. Figure 3 gives an example of using arrows from a Manual Handling course.
- **Video clips:** These are short clips (3 minutes maximum) and they are coupled with text and narrative to explain some important concept in a visual way.
- **Quizzes:** Come as multiple choice questions or True/False. Again, with every answer the trainee chooses comes an explanation.
- **Cyber tutor:** is virtual character that is used to comments on the learning material, explain particular points and emphasis on important issues. The cyber tutor can be seen as an alternative way to provide a trainer for the course. The trainee has the option to skip the cyber tutor section is they wish. Figure 4 shows an example of Cyber Tutor from Theft by Deception course.
- **Click to activate:** The trainee clicks on some points of interest to fire up extra explanation.
- **Combination of different interactive features together for larger scenarios.**

![Figure 3. Using arrows to show the right / wrong ways of carrying a load. Manual Handling Course.](image)

![Figure 4. Using Cyber Tutor to provide extra explanation and emphasis on some important points. Theft by Deception course.](image)

4. Evaluation Study

We conducted an evaluation study in 2016-2017 with employees of three large companies in the UK, aged between 21-69 years old. The sample size recruited was: Company A: 333, Company B: 597 and
Taking into consideration the type of the audience, the courses were designed to be short and interactive. The estimated time to finish the training part for the courses, including the interactive sections, was between 30 minutes to 50 minutes. Once a course became longer than 60 minutes, the chances of a trainee completing the course was reduced. For these training courses, the training part was independent from the test, and a trainee could take the test any time once they logged in. Figure 5 shows the average time taken to finish the training part by trainees for some of the courses. It is between 35 minutes, for Display Screen Equipment (DSE) course, and 45 minutes for Hijacking course.

To address our research question, we also logged the progress percentage as part of studying the performance of the trainees. Companies tended to concentrate on few courses that were of importance to them, and asked their employees to do those particular courses. Figure 6 shows the progress of finishing the training parts by the employees, where 100% means the employee finished the training part, 25% the employee went through only 25% of the training material, and so on. As it can be seen, more than 90% of the employees completely finished the training parts for the courses of their interest. We also conducted a subjective evaluation, results of which showed that the participants enjoyed interacting with the portal and 80% strongly agreed that it was user friendly.
5. Conclusion

In this paper, we presented T4L an interactive platform for offering training to industrial non-academic audience. The interactive methods used with these courses include drag and drop, using arrow buttons to change positions, short video clips, cyber tutor, and clickable points of interest. Using these interactive methods, a trainee has the chance to view the correct way and the incorrect way to perform tasks with short explanation of any chosen action.

Results of our evaluation study show that the average completion time of trainees was within this time range. In monitoring the performance of trainees, it was shown that more than 90% of the trainees completed the training part of the courses of their interest. As part of future work, we will further analyse the subjective evaluation questionnaire to find out what participants thought about different features of the portal. We will also look at personalising the content and presenting it based on user preferences and their interaction with the website.

References

Detect Students’ Academic Emotions in Classroom: Measurement, Self-perception and Manifested Behaviors

Yan Liu, Menghua Hu & Xiaoqing Gu *
Department of Educational Information Technology, East China Normal University, Shanghai, China
* xqgu@ses.ecnu.edu.cn

Abstract: The emotions students experience in the classroom are likely to influence their learning in school. For purposes of helping student improve their learning, this study seeks to further the development of the use of automated analysis of observed data to understand students’ emotions and during their learning activities. This study looks at linking behavioral data about emotions with self-report data about emotions with the intent to determine if observed behaviors are able to predict student’s emotions. The ICT-supported Learning Emotion Scale was used for 57 students to measure their learning emotion while learning in an ICT environment, meanwhile, classroom observations were conducted for 57 students for six weeks in 4 classrooms, paying close attention to student emotional behaviors and emotional changes. The results found that three key groups of behavioral indicators happened frequently. One of the key findings was that students who volunteered to answer questions in class were observed to have an observable “enjoy” emotion 84.79% of the time. The discussion highlighted that the bridge function of students’ emotion engagement between classroom observation and questionnaire data.

Keywords: Emotional engagement; Learning analytics; Behavioral indicators; Observations and questionnaire

1. Introduction

It is an ideal of teachers in classroom that the learning status of the students can be fully understood, in one way or other, thus the classroom teaching can be more adaptive and effective. This is what learning analytics is trying to accomplish, to develop a new way of “measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (LAK, 2011). While there are efforts of learning analytics focusing on different perspectives of learner behaviors and therefore in understanding learning status, students’ academic emotions during classroom session, are remain relatively untouched, due to the challenges of detecting the emotion data from the live classrooms. The state of academic emotions opens a window in understanding studying, therefore, attract lots of research on this topic, among which the constructs of academic emotions, the ways of measuring academic emotions, and the scales used to measure academic emotions are the main topics in this area.

From the point of learning analytics, measuring and understanding academic emotions as a potential way to unfold learning, should be manifested as measurable behaviors, thus that they can be collected and analyzed together with other learner behaviors and learning context information for the purpose of learning analytics (LAK, 2011). Although there have already been a lot studies on academic emotions, most of the methods are using self-report to measure students’ perceived emotions. The purpose of this study is therefore trying to propose a method to measure students’ academic emotions with the manifestations of students’ behaviors. By approximating students self-reported emotions status with their actual behaviors indicating their hidden emotions, this study aimed to provide a practical way for learning analytics in collecting important data of learner behaviors. Two steps were involved in carrying out this study, where first step was to adapt and verify a scale measuring students’ academic emotion with a self-report method; while the second step was to further verify the constructs and indicators of students’ academic emotions, by approximating the self-reported emotions with their actual behaviors.
2. Literature Review

Emotions are multifaceted phenomena involving affective, cognitive, physiological, motivational, and behavioral components (Damasio 2004). With the gradual recognition of emotions, emotions have gained a growing attention in education in recent years (Zembylas, 2007). Academic emotions refer to emotions related to students’ academic learning processes, mainly including enjoyment, pride, hope, anger, shame, boredom, anxiety, and hopelessness (Pekrun et al., 2007). In the last decade, numbers of scholars have focused on understanding the role of affect, or moods and emotions in education (Efklides & Volet, 2005). Previous studies have showed the important effect of academic emotions in students’ learning outcomes such as learning motivation, learning strategies, learning engagement, self-regulation, problem-solving, and academic achievement among others (Kim & Pekrun, 2014).

Emotional engagement is comprised of positive and negative affect in interactions with teachers, peers, schoolwork, and the school (Christenson & Reschly, 2012). Positive emotional engagement includes enjoyment, pride and hope, and negative emotional engagement includes anger, shame, anxiety, boredom (Pekrun et al., 2006). Previous studies have pointed out that emotions will indirectly affect students' academic achievement through engagement (Linnenbrink, 2007; Linnenbrink & Pintrich, 2004). González et al. (2015) also pointed out that behavior engagement is accompanied by emotions, such as enjoyment, enthusiasm, pride. While a large amount of research on student engagement during adolescence exists, with the ICT developed, there are some empty gaps during classes.

3. Methodology

This study is part of a learning analytics project to measure students’ emotion engagement through ICT-supported learning (iPads & PCs) in East China. A mixed-method design of both quantitative and qualitative techniques is employed, involving four classrooms in a Secondary school. Totally, we observed 57 students (31 female; age: M=12 years), who were a subset of students, four lectures in six weeks, including English, Chinese class.

The engagement only consists the emotion aspects in this article, Skinner once pointed out not the entire range of positive academic emotions are concluded in the emotional engagement (Meyer & Turner, 2002), but some energized emotional states, such as enjoyment, enthusiasm are included. However, when the students don’t engage in the learning task, they may behavior disengagement. The on-task behaviors and off-task behaviors were divided active initiative, working, and passive, which three categories can distinguish the students learning behaviors between positive emotion and negative emotion. Furthermore, the observers need to note each behavior with emotions (enjoy, bored, and anxiety). For example, one student was putting his hand in the classroom, while his mood was enjoyment. The instrument was measured in Table 1 based on the literatures.

The questionnaires were used to investigate the secondary students’ emotion toward ICT, which was developed from AEQ, and attempted to measure the students’ emotions during class. Those items including enjoyment (4 items), anxiety (5 items), shame (9 items), boredom (11 items). The study validated the 29-item by confirmatory factor analysis, presented with a five-point Likert scale, ranging from strongly agree (5 points) to strongly agree to strongly disagree.

There were two observers and their starting times were aligned. Beginning with the first student, the observer recorded the occurrence of any of the behaviors and then systematically rotated through the team under study in 30-sec intervals (Horn et al., 1986), marked all student behaviors seen within that time and then moves on to observe the next student. Each student was observed for a period of thirty seconds and his/her behaviors and emotion will be recorded before moving to the next student. After one cycle of students (10-13 students), the observer returned to the first student and began another cycle. In all, the maximum of cycles, which were scheduled, was sixteen in one classroom. Reliability estimates are conservative when compared to real-time observations when two observers code simultaneously. An inter-rater reliability between observers reached to 0.794 before they started to
conduct coding for the three class sessions, and then reached an inter-rater reliability of 0.875, which is accepted to be a good figure for observation protocols.

Table 1: the Emotional Engagement Observation Protocol (Sample)

<table>
<thead>
<tr>
<th>On-Task</th>
<th>Off-Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-Task Active Initiative</td>
<td>B1 rising hands; B2 asking a question; B3 taking notes; B37 raising your hand to answer questions by electronic tools; B38 asking questions by electronic tools; B39 taking notes by electronic tools;</td>
</tr>
<tr>
<td>Off-Task Active Initiative</td>
<td>B14 Interrupt teacher; B15 others; B62 playing games with iPad; B63 taking photos with iPad;</td>
</tr>
<tr>
<td>On-Task Working</td>
<td>B5 question answering; B6 reading; B7 doing activities; B11 taking lessons; B12 listening to teachers’ reviews; B17 discussing; B18 rolling play; B21 reading aloud; B41 listening to the radio; B42 reading the e-textbooks; B43 doing activities by electronic tools; B44 APP;</td>
</tr>
<tr>
<td>On-Task Passive</td>
<td>B24 watching peers doing tasks; B46 listening to the teacher using electronic tools; B26 Ah, Ah; B47 comment from teachers by electronic tools; B48 do not understand what teachers said</td>
</tr>
<tr>
<td>Off-Task Passive</td>
<td>B27 disturbing peers; B29 sleeping; B30 whispering to each other; B31 looking around; B33 laughing; B34 unwilling to participate in the activities</td>
</tr>
</tbody>
</table>

Before completing the questionnaires, the research advised by the teacher who has been teaching participant students to inform the ability for students’ respond. Then, respect students’ willing and volunteer to finish the search online during class. It took about 20 minutes to complete the whole survey. The 29-item (1 = strongly disagree, 5 = strongly agree) ICT-LES (parts of AEQ; Pekrun et al., 2002) during class was used to measure students’ levels of emotion scales. Barlett’s Test of Sphericity and KMO Test were applied on the data sets of ICT-LES (enjoyment, anxiety, shame, and boredom). We obtained KMO, moreover, it was found that items which factor loadings were less than 0.5 cross-loaded were gradually omitted, resulting in the removal of 4 items about anxiety, and shame emotion. Furthermore, it is acceptable for reliability analysis (enjoyment (α=0.912), anxiety (α=0.944), shame (α=0.967), and boredom (α=0.979).

4. Results and Discussion

4.1 Descriptive Statistics for the Emotion Engagement and Behavior Engagement

According to the result of ICT-supported emotion engagement scales, the mean value of each emotion engagement is enjoyment (M=3.79, SD=1.01), anxiety (M=3.22, SD=0.70), shame (M=3.07, SD=0.44) and boredom (M=3.37, SD=1.21). It points out that students show more positive emotion engagement in the ICT-supported learning environment. Students also have less negative emotion engagement when they learn with the support of ICT, like boredom, anxiety, but less likely to appear the emotion of shame.

According to the behavioral data, the percentage of on-task behavior averaged across all observations (80.7%) was within the range of previously reported estimates of on-task behavior in K-9 school students (90%-50%). Conversely, the frequency of off-task active initiative behaviors accounted for 19.3% of the total behaviors. Specifically, the frequency of on-task passive was 52.4%, the frequency of on-task working was 8.8%, and the frequency of on-task active initiative was 19.5%. In the present study, the two most common types of on-task passive behavior observed included: listening to teachers saying (B11), question answering (B5), and praising (B13). The two most common types of off-task active initiative behavior included: interrupting the teacher with a nonacademic issue (B14),
and unwilling to participate in the activities (B35). The most common type of on-task active initiative behavior contained putting up hands to answer questions (B1).

### 4.2 Detecting Behavioral Sequences Patterns

In order to investigate changes and influence in the students’ behaviors, the behaviors sequences patterns were performed with observations nested within classroom using Observation Scale. In this study, we used classroom observation protocol to get students behavioral sequences in the ICT-supported learning environment. Then we use the GSEQ 5.1 software to analysis students’ behaviors. The study shows the analysis results of the students’ behaviors sequences after adjusted residuals (in Figure 1). The connectivity of sequence has statistical significance (p < 0.05) if the Z-value of a sequence is greater than 1.96 (Bakeman & Gottman, 1997). The greater the value, the more significant the relationship between the two behaviors, and the more likely to promote the generation of the next behavior. Therefore, we can get those significant behavioral sequences, such as B1→B33, B1→B30, B1→B31, B35→B16, B35→B17, B33→B1, B5→B33, B5→B32, B5→B17, B12→B30, B13→B15, B15→B35, B16→B13, B24→B24, B24→B23, B21→B35, B21→B12, B21→B22, B23→B7, and B22→B1.

![Figure 1. Adjusted residuals table (Z-scores) of students’ behavioral sequences](image)

According to those significant behavioral sequences, we can draw students’ behavior prediction trajectory diagram in the classroom learning. It indicates that when one student raising his hands and answer questions, there are three possibilities for the next actions: students are likely to laugh at him, or whisper with others, or look at other people's reaction as group 1 of key behavioral indicators (see Figure 2). When a student is named to answer questions, other students may be doing little tricks, or discuss his answer, or ridicule him, as group 2 of key behavioral indicators in the Figure 2.

The paper shows that if students have the passive on-task behavior, then, it will lead to students in the off-task state, and eventually lead to students do not participate in classroom activities as group 3 of key behavioral indicators (in Figure 2). Students who read aloud with the teacher will also lead students to not participate in classroom activities. It finds that when a student was looking at someone else's task, he may have an On-Task action, which in turn causes him to finish classroom learning activity as group 4 of key behavioral indicators (in Figure 2).
4.3 Evaluation the Key Groups of Behavior Indicators by HMM

At the core of this approach is a hidden Markov model methodology that builds students’ behavior models from data collected by observations. A hidden Markov model (HMM) is a tool for representing probability distributions over sequences of observations. In HMM, the state is not directly visible, but the output, dependent on the state, is visible (Blunsom, 2004). In this study, based MATLAB coding, the HMM consists of hidden states that are not directly visible, and is governed by three sets of parameters: initial probability (behaviors state by observation), the transition probabilities (emotion state by observation) between states, and output probability matrix. The behaviors patterns (the output probability) associated with each state (Jeong, & Biswas, 2008). The transition probability associated with a link between two states indicates the likelihood of the behavior transitioning from the current state to the emotion state. For example, the HMM model states students rising hands would demonstrate a 92.7% likelihood of transitioning to enjoyment emotion. Likelihoods less than 10% were not represented, and there were 8 sequences in Table 2 below.

Table 2: the output probability

<table>
<thead>
<tr>
<th>No.</th>
<th>Behavior state</th>
<th>Emotion state</th>
<th>transition probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B21</td>
<td>E1</td>
<td>79.1%</td>
</tr>
<tr>
<td>2</td>
<td>B1</td>
<td>E1</td>
<td>92.7%</td>
</tr>
<tr>
<td>3</td>
<td>B11</td>
<td>E1</td>
<td>67.8%</td>
</tr>
<tr>
<td>4</td>
<td>B5</td>
<td>E1</td>
<td>82.1%</td>
</tr>
<tr>
<td>5</td>
<td>B17</td>
<td>E3</td>
<td>39.3%</td>
</tr>
<tr>
<td>6</td>
<td>B15</td>
<td>E1</td>
<td>79.6%</td>
</tr>
<tr>
<td>7</td>
<td>B3</td>
<td>E1</td>
<td>91.8%</td>
</tr>
<tr>
<td>8</td>
<td>B23</td>
<td>E1</td>
<td>76%</td>
</tr>
</tbody>
</table>

The study contains the relations between emotion self-report and the emotional observations, organized into behavioral observations through HMM. In the finding, total activity level, duration of anxiety emotion and boredom, has been affected by negative emotions, and has effect on the internal relations. The highly behaviors showed that if one student answering the question with enjoyment emotion, then he would return to on-task active initiative state (B5→B17, Z-score = 2.93) in Figure 3, however, he would turn to off-task state with passive emotion having a 17.9% stationary probability. Another example key behaviors conducted that one student was off-task state, and he returned on the on-task state (B15→B13, Z-score = 2.38) if the possibility of 79.6% was being enjoying. There was probability one student being on-on-task working state turn to do activities passively with enjoyment emotion with the possibility of 76%.
5. Conclusion and Limitations

The study conducted three groups of key behavioral indicators by a practical way for learning analytics. On the one hand, the paper verified the impact of interest between the emotions engagement by self-report and emotional observations. On the other hand, through the association with the behavior analysis, it can predict students’ behavior outcomes. The findings can help teachers concerned about the key behaviors of students, and adjust the learning strategy to be on the better emotional. In the future, the study will collect more observation data in different schools.

References

Authoring Tool: a Collaborative Web Tool for eBooks Creation

Dilson RABELO*, Ana Emilia OLIVEIRA*, Carla SPINILLOb, Aldrea RABELOa, Rômulo FRANÇAa & Elza MONIERa

a UNA-SUS, UFMA, BRAZIL
b UFPR, BRAZIL
* dilsonrabelo.unasus@gmail.com

Abstract: The educational system is constantly changing mostly due to new technologies incorporated during the teaching and learning process. An example of these new educational technologies is the eBook. This tool has been considered of great importance, especially in the e-learning modality. Its development process may require a long time and a multi-professional team. Because of this, the production of quality eBooks may be discouraged among education professionals who do not master the necessary technological skills. Considering this reality, this paper presents the Authoring Tool (AT) as an alternative to optimize the process of eBooks creation for non-programmers. To accomplish it the tool uses a WYSIWYG interface. To test the tool's effectiveness in saving time during the eBooks production process, we will run initial tests. The results of these tests will be presented in this paper.

Keywords: eBook. Education. Authoring tool.

1. Introduction

The use of Information and Communication Technologies (ICT) has radically transformed the way people communicate, conduct business, do banking transactions, etc. These transformations have also reached the pedagogical paradigm, that is, they have altered the way people teach and learn and, consequently, how educational materials are developed and offered to the public (Audino & Nascimento, 2012).

In other words, with the emergence of technological resources in education, the teaching-learning process has gained the support to optimize its processes and achieve the practicality and efficiency necessary to promote an interrelation with the students during their learning. Among these supports there are the digital learning objects that have emerged in recent years as internationally recognized elements for the construction and publication of educational contents (Silva, Silva & Guedes, 2014). In this context, the eBook stands out as a book in digital format that can be used as a resource for Distance Education courses. However, although it is a widely used resource, its creation requires considerable time and effort. In order to improve the efficiency in the creation of these eBooks, we planned the development of an authoring tool.

This article presents a web authoring tool for the optimization of the eBooks development process. It will also be presented its authorial process, the functionalities, a pre-experiment to evaluate the tool’s efficiency and a workshop evaluation of the tool with designers.

This paper is organized as follows: in Section 2, the model of eBooks adopted by the tool is presented. In Section 3, the tool's functionalities are explored, with details of the pedagogical design stages and the transformation in the learning object (LO). In Section 4, the result of the pre-experiment performed on the tool and the outcomes of the workshop evaluation are explained. And finally, in Section 5, there is a discussion of the partial results and the next steps for the improvement of the tool.

2. Ebook Template

With the advent of new technologies, education has undergone changes in the way of representing its learning objects, seeking to keep pace with technological evolution. The printed book, for example, has
been improved and adapted to a new format that accompanies this evolution. This new format, called eBook, besides presenting texts and images, has new elements such as hyperlinks and multimedia objects (Seadle, Vassiliou & Rowley, 2008).

The focus of the eBook is to allow the reader a more engaging and interactive reading experience with the content presented (Dziekaniak, Moraes, Medeiros & Ramos, 2010). Thus the eBook template used in the Authoring Tool was made taking into account some pedagogical aspects (Reategui & Finco, 2010).

Figure 1 shows an example of the eBook. It was developed using a set of technologies: HTML5 (Developed by MIT, ERCIM, Keio, Beihang, 2014), CSS3 (Developed by MIT, ERCIM, KEIO, 2011) and Javascript (Developed by MIT, ERCIM, Keio, Beihang, 2014). The informational design was organized with well-defined elements, in order to facilitate the navigation and content visualization. Some interface criteria involved in its development are described below, following the model defined by Reategui & Finco (2010):

- Use of images: according to the principle of multiple representation, it is better to represent an explanation with texts and illustrations than with texts only;
- Presentation of texts: it is important to present the texts in an appropriate way, observing some aspects such as contrast between the fonts and the background of the screen, facilitating the reading;
- Orientation and navigation: to allow the user to locate in the presented resource, visualizing what was done and what is available;
- Interactivity: the user must interact with the learning object, being able to have a range of possibilities during its handling;
- Aesthetics: the learning object must have features that make it pleasant in its visualization;
- Affectivity: to allow the resources presented in the learning object to express their affective states, such as a movement of images.

Figure 1. Template of the eBook following the principles of Reategui: (1 and 2) orientation and navigation, (3) images, texts, interactivity, aesthetics, and affectivity.

The eBook page is divided into header and content. The header is where the navigation menu, configuration and information menu are located. Page navigation is the area where the reader can access the pages. In the setup menu, the reader can choose the font size and access the summary to select units.

With the popularization of mobile devices such as smartphones and tablets, a learning object with a mobile web interface should allow the same page to be displayed correctly by all types of devices. A simple solution that fulfills this requirement is an interface based on Responsive Web Design (Baturay & Birtane, 2013). Figure 2 shows how the eBook is displayed on screens of different resolutions.
3. The Authoring Tool

The eBooks development process, explained in Section 2, requires the integration of three teams: IT (Information Technology), ID (Instructional Design), and GD (Graphic Design).

At the beginning of the process, ID team elaborates the pedagogical part and a prototype of the eBook. In parallel, they request DG team the graphic materials that will compose the book. With the prototype and graphical features ready, this artifact is sent to IT that initiates the development of the eBook. At the end of creation, IT forwards the book for validation. Once approved, it is made available to the VLE (Virtual Learning Environment).

Based on this scenario, it was proposed the creation of an authoring tool that would aid in the optimization of eBooks development process: the Authoring Tool.

The eBooks authoring process in the tool has two actors responsible for the conception of the learning object, the designer, and the layout editor, who operate in five functionalities during the process. The eBook authoring flow can be summarized as shown in Figure 3:

![Figure 3. Authoring flow.](image)

3.1 Project

The project is the beginning of the eBook authoring process. The designer should be concerned with the definition of three variables in the creation of the project. The first is the definition of the data that make up the project, the second is the definition of the layout editors and, finally, the establishment of the structure of the eBook.

In the eBook structure, the designer defines the module(s), unit(s) and chapter(s) that will compose the project. Each of these items can be edited, deleted, or moved by the designer. Pages are created next to chapters during the authoring process, in the Canvas step.

3.2 Theme

The themes are the definitions of the templates containing the characteristics that will be presented in
the eBook, such as colors, font size, and spacing, among others. The themes are represented through a script in CSS3 format, being injected into the export process.

3.3 Media

The tool has a media library. It offers support for JPEG (Joint Photographic Experts Group), SVG (Scalable Vector Graphics) and PNG (Portable Network Graphics) images. Another feature in the library is the ability to associate tags with images, to classify the elements that make up the image bank. In the current version of the library, the upload of videos is not supported yet.

3.4 Canvas

One of the most prominent features in the Authoring Tool is the Canvas area. This feature is the working area of the layout editor. The interface follows the WYSIWYG principle, allowing the user to directly manipulate eBook pages with clear feedback on what the final presentation will look like. Figure 4 shows the Canvas interface.

![Figure 4. Canvas Interface.](image)

Canvas has a palette of components (1) grouped by types: layout, text, effects, and media. Each group has a range of components that can be added and manipulated on the canvas (3). The toolbar (2), provides the user with some shortcuts, for example: undo, redo, and manipulate book structure, among others. Together with the canvas, the users obtain auxiliary navigational resources (4), which allow them to save, preview the project, export and access the project media.

The project structure is the area where the layout editor will manage the pages, and then define the content that will be part of the eBook.

3.5 Export Process

The export process is the end of the authoring process in the tool. In this step, the user can export the eBook developed. Currently, the export format supported by the tool is HTML5 (W3C, 2014) along with Javascript and CSS3 (W3C, 2001). The export process is executed in 3 phases: First, the CSS of the template is injected into the eBook; second, all pages undergo a cleaning process, executed by a Garbage Collector, contained within the tool, and finally, the book is compressed in zip format and made available for download.

4. Results of the Tool’s Evaluation

4.1 Pre-experiment: Partial Results

We attempted to evaluate the time for development of an eBook created directly in the authoring tool. The focus was to identify the work evolution. For the experiment, developers in the technology field with Web development experience were invited to participate in the test. They were in an assisted environment, with the same conditions usually established for the development of an eBook with 30
In the common production scenario, without using the Authoring Tool, it took an average of 8 hours to complete the task. The development of the eBook with the Authoring Tool proved to be promising as developers were able to finalize the same material in half the time. In this initial sample it is possible to highlight positive and negative points.

As negative points of the process, it was observed:

- The eBook follows a system of grids, rows, and columns to position its elements. During the authoring process, the participants had difficulties in handling with the elements of the grids;
- Because it is a version in development, some problems interfered in the eBook authoring process.

As positive points, it is worth mentioning:

- The tool encapsulates the complexity of creating the components. This factor was crucial for accelerating the development of the eBook;
- The components grouped into menu allow easy location and inclusion in the canvas.

The results obtained with the experiment show a promising scenario, since it was observed that the tool helped to optimize the production of the eBook. It is clear that more experiments should be applied in other scenarios to validate the proposal.

4.2 Outcomes of the Workshop Evaluation

In order to assess the information design aspects of the Authoring Tool (e.g., user interface, interactive graphic resources), a workshop evaluation was conducted with 10 participants with expertise in this field. Initially, an explanation of the AT functions was provided to the participants through a live tutorial. Afterwards, participants were asked to design an eBook with any content of their choice. At the end of the workshop, each participant presented their eBooks to the group. Then, they answered an online questionnaire to evaluate the tool on its ease and difficulty of use and to give suggestions for improvements. The questionnaire resulted in a total of 109 responses, which were compiled into the following categories: (a) Perception of Experience, (b) System Design, (c) Interface Design, and (d) Task Execution (Table 1).

Table 1: General results of the questionnaire.

<table>
<thead>
<tr>
<th>Category</th>
<th>Facilities</th>
<th>Difficulties</th>
<th>Suggestions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception of Experience</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>System Design</td>
<td>6</td>
<td>19</td>
<td>17*</td>
<td>42</td>
</tr>
<tr>
<td>Interface Design</td>
<td>5</td>
<td>15</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>Task Execution</td>
<td>14</td>
<td>9</td>
<td>-*</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>46</td>
<td>29</td>
<td>109</td>
</tr>
</tbody>
</table>

Number of responses per category:

Less than (<) 10: respondent did not express an opinion in the category
More than (> 10): more than one respondent expressed an opinion in the category
(*) For compiling the results on suggestions, the categories System Design and Task Executions were integrated, as it was not possible to differentiate the participants’ responses on them.

The responses on Perception of Experience consider participants’ overall impression of their experience with the AT during the workshop. These were expressed through opinions of an emotional nature and/or value judgment. A total of 16 responses fit into this category. In general, participants found the AT easy to use (N=9), considering it an innovative and user-friendly tool. Participants also considered the tool functions and features easy to use when carrying out tasks to design their eBooks during the workshop. This was shown by the 14 positive responses out of 23 in the Task Execution category, among them, the inserting of rows, columns, graphics, images, videos, pages and texts.

On the other hand, most responses on the System Design category (N=19 out of 42) were related to participants’ difficulties in using the tool. These concerned features or functions that were not available or limited the users’/participants’ actions when interacting with the AT. Statements about restrictions on typographic features and choices (e.g., typefaces, text alignment, font size, color), and
the excessive number of actions to insert an image are examples of responses found in the questionnaire. Similarly, most responses on the Interface Design category (N= 15 out of 28) reveal participants’ difficulties, due to problems in the visualization of elements and/or functions in the AT interface. Inappropriate labelling of buttons, unclear icon designs, lack of visual/graphic cues to indicate the selection of an element in the interface are examples of the participants’ responses in this category.

Finally, a total of 29 suggestions to improve the AT were made by the participants. These suggestions mainly relate to the System Design and Task Execution categories, as for instance to include floating menus or shortcuts to menus, to add image editing tool (e.g., resize, crop), and to enable the categorization of images by an eBook project. The suggestions were then, organized on a checklist addressed to the AT developing team so that they could make the adjustments to the tool.

5. Conclusions and Follow up

This paper presents the Authoring Tool to develop eBooks with the purpose of assisting in the optimization in the development of these learning objects. Still in its beta version, it was possible to perceive a promising future for the use of the technology. Based on the experiment, a 50% gain in eBook production time was observed, a satisfactory but not conclusive result.

Despite the promising results regarding the optimization of the eBooks creation time with the Authoring Tool, more experiments will be conducted in order to identify successes and potential problems, and, based on these results, follow with the development of the tool to supplement the proposal described in this article. In this sense, the outputs of the evaluation workshop also contributed to improve the AT with regard to information design aspects. A revised version of the AT with a user centered design approach is currently being developed for further evaluation.

Acknowledgements

Thanks are due to the participants of the evaluation studies presented in this article, and to the Brazilian Ministry of Health, and to FAPEMA-PRONEM for funding the development of the Authoring Tool and the related research project.

References

Observing the Degree of Distortion in Coordinated Motor Actions

Takehiko YOSHIKAWA, Kohta SUGAWARA, Kenji MATSUURA*, Karungaru STEPHEN & Naka GOTODA

Graduate School of Advanced Technology and Science, Tokushima University, Japan
Information Technology Center, Kagawa University, Japan
*ma2@tokushima-u.ac.jp

Abstract: Integrative motor actions have a characteristic that consists of multiple primitive motions. A time-series of movement includes a unit of integrative motor actions. Movement rhythm arises from coordinated efforts in this exercise. This study tries to support learning, focusing on double-under in rope-skipping. Double-under has major primitive motions: jumping and rotating wrists. There is timing difference between primitive motions. Thus, we design and develop a supporting system for learning timing difference in dynamic sense. The system measures learner’s data with wearable sensors, calculates the appropriate time and suggests feedback.

Keywords: Motor skill, repetitive exercise, learning support, wearable sensor

1. Introduction

This study focuses on learning support for motor skill based on human embodiment. It deals with motor skills as an ability obtained through learning in order to execute tasks requiring body motion. In particular, the domain on this study is continuous skills that are repetitive such as rope-skipping or hula-hoop. Among them, the target of this study is double-under skill in rope-skipping. The characteristic of double-under in repetitive exercise is to repeat movement pattern that is divided by a time. Fujinami (2006) states a rhythm arises from cooperative motion. This study regards it as rhythm of double-under and especially focuses on timing that is one of the timewise elements of skill. Johnson (1961) states the dimensions of skill have been identified as speed, accuracy, form and adaptability. Accuracy include skill done in good timing. To execute double-under, it is necessary to control primitive motions cooperatively at the right timing.

Besides, double-under is a sort of an integrative exercise with primitive motions classified broadly into two, such as jumping and rotating wrists. Each primitive motion usually has unique timing. There is a time difference between them. Those cooperative motions should have a certain association such as interval of each timing. The more fixed timing difference becomes, the more rhythmic double-under is. Thus, a learner’s goal is an improvement of timing difference stability. However, as one of an issue with this improvement, cooperative motions require a skill for it. We design a learning support so that this difficulty is relieved. Therefore, we propose a learning support system in order for a learner to acquire an appropriate timing leading rhythmical movement.

2. Preliminary Discussion

2.1 Timing on Double-under

This study regards timing as one of the essential components of rhythm. In rope-skipping skills, double-under is more influenced by the feature of ballistic exercise. This exercise needs to calculate an expected time of next movement. The movement pattern on double-under becomes a segment that is not modified as ballistic exercise. In addition, double-under is an exercise requiring open skill. While performing it, we need to observe a trajectory of the rope because the timing differs at each time.
This study defines the appropriate timing of primitive motion on double-under as follows. Timing of rotating wrists in this study is the starting time at the first rotation to begin the rope movements (Karungaru et al., 2016). The actual rotation during a jump is twice per jump. The present study picks out the first time as timing of rotating wrists because it is a sort of ballistic exercise. Jump timing is defined as the starting moment of jumping. Figure 1 shows the image of the actual timing of primitive motion in double-under. Difference between jump and rotating wrists is not the motion but the phase difference between each wave form. This study defines the phase difference as a gap of timing difference by a feature of double-under for integrative exercise. We regard a mandatory task of the skill integrating motions as an inhibiting reason for learners from acquiring double-under. The degree of distortion in waveform makes observed data vary depending on the skill.

Figure 1. The timing of primitive motions

2.2 Theoretical discussion for Learning Support

Fitts and Posner (1967) classify learning stages into three. In the second stage, learners organize more effective movement pattern because they obtain embodiment to some extent in motor stage. Until the last stage at which they can acquire unconscious movements, they need to pay attention to learning the movement way. In particular, in motor stage, learners monitor feedback in order to improve their own performance gradually. On the conceptual model of human performance, it has two major categories of feedback: intrinsic feedback and extrinsic feedback (Schmidt, 1991). Intrinsic feedback is the information from visual sense, auditory sense, proprioception and so forth. On the other hand, extrinsic feedback includes a system that supports learners. It has two important information: knowledge of result (KR) and knowledge of performance (KP). KR is the information on whether movement is success or not. When we want to give more concrete advice, it means that the feedback must have kinematic information.

Besides, the present study assumes that movement differs from both body form and performance thereof every time. In relation to this, Schmidt (1975) states that motor-learning can be generalized with a schema theory from a skill science viewpoint. According to this theory, learners acquire schema that is rule-based. The rule is related to the joint result of trials itself and associated parameters for realization. Learners strengthen and renew this rule through practicing overtimes. Our target is learners who are in motor stage. They refine the schema derived from the past information stored inside, which is an experience. Each parameter that is able to succeed on past trials can comprise schema as with accurate parameter setting. Our support suggests a criterion for comparing with intrinsic feedback. Learners revise the error between this suggestion with feedback and learn so that they can decrease it.

2.3 Requirements

In this study, timing difference of movements between two body parts is observed. It detects elongation and contraction of the time scale identified as relative criterion. The anchoring point in this study is set to the highest time of the body movement, that is called jump timing. We support relative timing of rotating wrists against the anchoring point. The system observes two body parts in parallel for the sake of the detection abovementioned. It monitors myogenic potential of the forearm because muscle contraction causes the movement. On the other hand, it identifies jump timing by monitoring the acceleration of the waist.

Real time support seems better when learners try to stabilize double-under movement requiring open skill as the best support timing depends on current condition. When the system suggests timing as feedback to learners, it is better that there is not different between suggestion and actual motion timing. According to the literature review (Choshi, 1972), auditory sense works for a quick response to external
stimulation during 120-185 milliseconds and haptic response does during 115-190 milliseconds while visual response does during 150-225 milliseconds. As a synchronous support, it is necessary to choose suitable feedback media. Moreover, the system needs to have two models for the support; i.e. a predictive model and a supporting model. By comparing current data to this model, it tries to realize the prediction. After prediction, the model for supporting helps it to determine the time showing feedback. This study utilizes the data on past trials in each person for determining each parameter as the models just like as schema is shaped.

Furthermore, in case of monitoring double-under by a wearable system, a system should avoid wired communication with sensors because it is more restrictive. In addition, the size and the weight of each sensor are problematic. Due to these considerations, this study adopts a microcomputer that is small and light. Additional advantage of the microcomputer is to combine the role of multiple inputs by sensors and an output to show feedback as an actuator. Therefore, a microcomputer linked to sensors and an actuator should be small and light under the condition that it is free and not an obstacle to performance.

3. Prototype System

We designed and developed a prototype system that has three basic functions: observation, analysis and feedback. What we had in mind in the initial design has been published (Yoshikawa, 2016). The total system weight is about 120 grams. A learner attaches the system consisting a microcomputer and an acceleration sensor to her/his own waist using a belt. This attachment method and weight are within a range of ignorable influence on embodiment.

Figure 2 shows the flow of the system support. The system needs to prepare models for real time support. Before creating the models, we define a local maximum on acceleration data as jump timing. We utilize average rectified value (ARV) for defining timing of rotating wrists. For computing ARV, this study regards myogenic potential during rest as 0, and rectifies it by a method called full-wave rectification. After that, low-pass filter with a time constant of 30 milliseconds smooths the rectified wave. Then, the system determines local minimum in ARV wave as timing of rotating wrists. However, raw data sometimes mixes with action potential leakage of adjacent muscles called crosstalk. When this case or becoming dull by smoothing happens to the waveform, we regard extremum values in differential values of ARV as timing of rotating wrists. Based on them, we decide each parameter in the models. Jump timing arises from double-under of repetitive exercise that closes the same cycle. This study treats it as a fixed value. There is timing difference between jump and rotating wrists. The learner's goal is to induce learner’s diminution of dispersion on timing difference. Thus, the parameter on this system is the fixed value suitable for supporting as timing difference. When we set parameter as noted above, the system is ready to support.

![Figure 2. Flow of support by using this system](image)

On real time support, at first, the system measures a learner’s performance by monitoring myogenic potential of the forearm and acceleration of the waist. An electrode lead needs to be immovable on learner’s body so as not to disturb the motion of rotating wrists because artifact gets mixed in the data. The microcomputer to control these sensors is Arduino Fio having a XBee socket.
that enables it to communicate wirelessly. Another computer is set up in order to store the data that is received from the system through the wireless network. The system observes a learner’s performance every 10 milliseconds. Secondly, the microcomputer analyzes performance data. Synchronous support system has to identify the timing and predict learner's movement within acceptable time range. The microcomputer executes the prediction that synchronizes with movement. Finally, the system provides feedback to learners as function. The supporting content is the first timing of rotating wrists in each jump. The time lag happens even if the system stimulates auditory sense. Hence, we suggest a sound of the primary timing on support 120 milliseconds earlier than the ideal one as a revision of the time lag in reaction to support. The system provides feedback by a speaker controlled by Arduino Pro Mini that is a different microcomputer from the Arduino Fio. After completion of monitoring learner’s embodiment, the system updates the parameter using monitored and stored data. In this way, the more learners use the system, the more suitable parameter the system assigns.

Learners get learning supporting context by training double-under while equipping the system. Since the support content is the first timing of rotating wrists in each jump, they receive one sound from the system every one cycle. Although learners’ goal is learning the relative timing, the system relies on them to perceive criterion timing by intrinsic feedback. They solve this problem by training while recognizing timing different. Thus, they adjust motion timing to the suggestion gradually.

4. System Design and Development

4.1 Overview

The computer, which communicates with the wearable system using wireless communication, is only needed to store data received from the system in terms of the basic design of the prototype. The computer has powerful resources compared with the system. The analysis function in this system can move to the computer because it does not have to respond so quickly. Other than this, input/output functions are still necessary on the wearable system side; i.e. the part of myoelectric sensor. Ongoing prototype sensor system sometimes has the risk of mixing data and noise. The current design tackles to integrate Myo (https://www.myo.com/) for monitoring myogenic potential. It has 9-axis inertial measurement unit besides 8 separate electromyography (EMG) sensors. Those sensors attached to the forearm help this system more comprehensively figure out what the wrist is doing. It can send measured data to the computer using Bluetooth and reduce the risk. When the system converts the data with less noise into ARV, this waveform is useful to understand muscular activity and support learning double-under skill.

However, this device doesn’t get acceleration data for monitoring jumping. When the computer analyzes performance data, the system doesn’t need high performance microcomputer like Arduino Fio. It only has the functions of sensing and transmitting wirelessly. Hence, we can reduce its size and weight by the system measuring acceleration in parallel to control a speaker. Learners attach a speaker and a sensor to their body together. An attachment region of the microcomputer is at instep or ankle because it is more direct for detecting timing when the system measures the position of the foot.

The major improvement of the system is the analysis by the remote computer. Learners often have the distortion in movement cycle like Figure 3. Thus, the system needs to predict next motion dynamically.

![Figure 3. The distortion in jump cycle](image-url)
4.2 Consideration for Dynamic Prediction

In our belief, novices perform double-under in unsystematic movement because they need embodiment with open skill just as jumping depends on motion of the rope. Thus, synchronous support for learners that have dispersion of jumping cycle expects the next timing dynamically from a current state. Jumping especially has a feature of repetitive exercise. Focusing on the centroid, this is simple vertical movements. Figure 4 shows the graph transformed time-series data of acceleration into time-delay coordinate system. The symbol \( y(t) \) is the acceleration value of the waist at time \( t \). Thus, the x-axis shows the value of acceleration at the current time and the y-axis shows the acceleration value at the last time. The shape converges to an ellipse shape to some extent.

When the system follows itself, one of the predictive methods is to draw approximate curve in the neighborhood in Figure 4. The graph line appears along with an approximate curve because behaviors seems to be similar to the past. This method may become more accurate if we measure the data by reducing noise. Besides, regression analysis seems useful for implementing dynamic prediction. Toyooka et al. (2016) propose a system that estimates the next cycle on repetitive exercise using regression analysis. This study needs to calculate the regression formula from any suitable data on double-under performance. It has some analogy with the above method using an approximate curve. Thus, we look for two kinds of data group that have relation to regression analysis.

![Figure 4](image-url). The acceleration data in double-under transformed to time-delay coordinate system

4.3 Feedback for Effective Learning

The influence of extrinsic feedback to learning is inestimable. Nevertheless, the prototype system doesn’t provide KR. In case the system only gives sound to learners, we shouldn’t compel its feedback to include KR. In spite of getting feedback once each cycle, learners feel that they cannot stop concentrating their attention on either embodiment or sound. We suppose the reason is too abstract feedback. There is characteristic sound that is wind noise arising from reiteration of movement pattern in double-under. This system provides sound three times a cycle as rhythm of double-under but we regard this pattern as once ballistic motion. The reason is that we should try to grasp rhythm as the flow (Fujinami, 2006). In addition, Sugawara et al. (2016) state that double-under has the correlation between components of motion. On integrative exercise, we assume that the system should support either one, if one of the component in primitive motion developed with another one. Therefore, it is necessary to investigate what the correlation between timing and the others is.

The beginning of a segment is the time that ballistic motion begins. When our body executes planned movement pattern by the rhythm provided, it will be an indication of the next embodiment as illustrations of more concrete motion. Thus, the system shows intuitive feedback as KP using sound. In addition, Learners may require that the system gives KR. This system has a function of showing KR because this feedback lacks efficiently. Fortunately, Myo has feedback method that is different from sound. The time of human reaction is relatively fast because its vibration stimulates haptic sense. Myo provides KR simultaneously or separately with sound. KR is the result compared with embodiment and suggestion. We must be careful of interference between feedback.

4.4 Implementations
The system follows the condition that the learners obtain calculated movement. They learn an appropriate movement to fit the beginning of motion close to the suggested timing but we assume that they perform the motion after listening to the sound. However, they might feel that it is hard to use it. Thus, we need to reconsider the necessity of the refinement. In repetitive exercise, there is a possibility of unnecessary feedback because the rhythm arises from almost constant movement that is readily understandable. When the prototype system applies these improvements, it needs to change the value per individual and a degree of calculating.

Learners use the system adjusted in this way. The system also follows the supporting flow in Figure 2. They only equip it and jump. As the system suggestion is the criterion for becoming stable of timing difference, they should endeavor to tune their motions to it while training. The system suggests more feedback for acquiring rhythmical movements. Learners don't need to pay attention to timing difference while training it. We expect their learning to have a good effect on an acquirement of another repetitive exercise and so forth.

5. Conclusion

In this study, a learning support system on double-under as integrative and repetitive exercise is proposed. In particular, the present study focuses on timing difference of primitive motions in integrative exercise and the timing that movement rhythm arises from repeating the movement pattern in repetitive exercise. We identify them by observing acceleration of jumping and myogenic potential of rotating wrists. In addition, this study designed the new system function. The proposed system observes the data by wearable sensors, analyzes to predict movement dynamically and shows feedback founded on analyzed data. Although this paper designed the learning support system for double-under in repetitive exercise, we plan to carry out an evaluation on an effect of supporting by dynamic prediction. We try to conduct an experiment for proving an efficacy of the proposal. We also proposed an actuator using two channels: sound and vibration. Therefore, we must implement the system concretely for a practice as a future work.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP15K01072.

References

Learning Environment for Recursive Functions by Visualization of Execution Process

Raiya YAMAMOTO*, Yasuhiro ANZAI, Satoru KOGURE, Yasuhiro NOGUCHI, Koichi YAMASHITA, Tatsuhiro KONISHI & Yukihiro ITOH

*Graduate School of Science and Technology, Shizuoka University, Japan
bFaculty of Informatics, Shizuoka University, Japan
cFaculty of Business Administration, Tokoha University, Japan
dShizuoka University, Japan

*yamamoto.raiya.15@shizuoka.ac.jp

Abstract: In programming learning, it is difficult to imagine the behavior of a recursive function. This difficulty stems from the following: (1) learners have to manage different instances of functions with the same name, (2) they cannot understand the execution process of passive flow, and (3) novice learners cannot manage unnecessary portions as a black box. Furthermore, it is desirable for authors of teaching materials to customize the visualization of each instance individually. In this study, we have extended a program visualization tool (TEDViT) to be able to visualize recursive functions by satisfying the above issues. We conducted an evaluation of our method and obtained positive results.

Keywords: Visualization, programming education, recursive function

1. Introduction

In programming learning, it is thought to be difficult to imagine the behavior of a recursive function. We believe that this difficulty exists due to the following three reasons: (1) learners must manage different instances of a function, (2) they cannot understand the execution process of passive flow, and (3) novice learners are unable to manage unnecessary portions as a black box.

For (1), a recursive function calls the same function recursively; therefore, learners need to manage different instances of the same function. It is difficult to reproduce the behavior of each instance and to retain these behaviors in their brains, because these instances are produced from the same function when recursive calls are executed. To address this, we consider it is necessary to visualize the execution processes of instances with a bird's-eye view (Common Requirement 1).

Regarding (2), Sholtz et al. revealed the source of the difficulty of imagining a recursive function’s execution process: the control flow, known as “passive flow” (Scholtz, et al., 2010). In a recursive function, a new instance of the function is called, with control repeatedly passing forward to the newly created instance (active flow). Then, the flow reaches a termination condition of the recursive calling, after which the control and a value are repeatedly returned to the calling function. This flow is known as the passive flow. Scholtz et al. mention that the difficulty of the passive flow stems from novice learners misunderstanding that execution process of a recursive function ends immediately when termination conditions are satisfied; Shortly, they never imagine passive flow. To address this, we consider it is necessary to visualize behaviors by suggesting passive flow (Common Requirement 2).

In terms of (3), a recursive function’s execution process tends to be complex, due to increasing instances of the same function, and the need for learners to recognize passive flow. To prevent this, a method for hiding the called functions’ behaviors and simplifying the entire process is required, which we refer to as “black boxing.” To achieve black boxing, we consider it is necessary to visualize behaviors using a situation in which called functions’ behaviors are concealed and they only return output (Common Requirement 3).

Furthermore, it is desirable to customize the visualization of the behaviors of each instance of recursive functions individually, because data structures differ in program algorithms and the data
managed and received by called functions vary. For this reason, it is preferable for authors of teaching materials to be able to customize the ways of visualization (Individual Requirement).

Various tools have been developed to visualize programs’ behavior with animation. Jeliot3 (Moreno, et al., 2004) is a tool for visualizing Java program behaviors. However, this tool does not satisfy the Individual Requirement and Common Requirements 1 and 3. SRec (Velázquez-Iturbide, et al., 2008) is a visualization tool to assist understanding recursive functions. However, this tool does not satisfy Common Requirement 3 and the Individual Requirement. ANIMAL (Rößling and Freisleben, 2002) and TEDViT (Kogure, et al., 2014) are tools to visualize program behavior using animation. Teachers can setup the visualization with scripts or rules, so they satisfy the Individual Requirement. Moreover, teachers can write scripts and rules that satisfy the Common Requirements. However, these scripts and rules are not specialized for recursive functions, so a great deal of time is required to write scripts and rules that satisfy the Common Requirements. The Common Requirements are common features for managing recursive functions. Hence, these features should not be written with scripts and rules, but rather embedded in systems. In comparing TEDViT and ANIMAL, the cost of writing TEDViT rules is less than that of ANIMAL scripts, so we extend TEDViT.

Consequently, the purpose of our research is to extend TEDViT to be capable of managing the Common Requirements for recursive function visualization. In this paper, we report on our process of extending TEDViT, and the evaluation thereof.

2. Supplemental Features to Apply Learning Assistance for Recursive Functions

2.1 Basic Function of TEDViT

Teachers can customize the visualization of unique algorithm concepts and program behaviors with TEDViT (The interface is shown in Figure 3). The figures are reproduced in an area called the status of target world (STW). To use TEDViT, teachers should write rule sets (hereinafter, “drawing rules”). Drawing rules allow TEDViT to display data objects with effects. TEDViT supports objects: variables, arrays, arrows indicating relationships between data, and so on, as well as visual effects: colors representing emphasized parts and balloons showing explanatory texts, among others. Learners can trace program steps with these visualizations.

2.2 Visualization of Instances of the Same Function

As mentioned in the introduction, recursive functions call themselves recursively and create instances from the same function, so making it difficult for learners to manage their behaviors using only their brains. For this reason, TEDViT should show the functions comprehensively, and assist learners in focusing on a particular function. However, TEDViT currently can visualize only one function at a time. Therefore, TEDViT needs to be expanded to be able to generate a new drawing area on the STW and visualize objects corresponding to variables, for example, on the generated drawing area when recursive calling is executed. Moreover, to know which recursive function calls a function recursively, TEDViT needs to visualize drawing areas as a tree structure so that learners can see which area is generated recursively and can know their calling/called relations. We refer to the drawing area as a “function drawing area” (hereinafter, “F-Area”). An example of the F-Area is shown in Figure 3.

2.3 Visualization of Passive Flow

As mentioned in the introduction, it is difficult for learners to recognize passive flow, as they misunderstand recursive calling ending when the termination condition is reached. Furthermore, in most cases, passive flow exhibits a complex process. Therefore, it can be challenging to imagine the passive flow process of data returning to the calling function correctly. For this reason, TEDViT needs a new visualization method so that learners can understand the passive flow process.

As mentioned in 2.2, the extended TEDViT can visualize F-Areas by connecting them in a tree structure, and learners can thereby see the data flow among functions. Figure 1 demonstrates how the passive flow is observed. By observing behaviors from the structure, learners can check the values that
the calling function obtains as the output of the called functions, and understand the flow of returning output of the called functions. We refer to these functions visualized as a tree structure the “recursive function tree.” However, learners may not realize the existence of passive flow using the recursive function tree alone. The solution to managing this problem is described in 2.5.

![Figure 1. Observation of Passive Flow](image)

2.4 Visualizing Recursive Calling with Black Boxing

When writing recursive functions, we don’t imagine the called function behaviors, but code the process considering the return values obtained from the called functions. By black boxing calls to subordinate functions, we can hide and simplify complex function calls, and observe an overview of the recursive function execution process. However, it should be recognized that black boxed called functions behave in the same manner as calling functions. If function calls are black boxed and we do not consider concrete behavior, this means hiding recursive features and termination conditions. For this reason, in order to visualize recursive calling effectively, the following requirements should be satisfied; The ability to observe recursive calls by means of their being black boxed (Requirement for black boxing 1). The ability to observe recursive calls by disabling black boxing from the top to bottom function (Requirement for black boxing 2). The ability to enable learners to consider the termination condition (Requirement for black boxing 3).

To achieve requirement for black boxing 1, we add the process that behaviors of called functions is hidden and the behaviors are skipped until the called function returns a value (hereinafter, “1-step recursive calling”, the left side of Figure 2). When subordinate functions are called, the behaviors of the called functions are black boxed and learners observe only results from the called functions. Learners also can disable black boxing by clicking the black boxed F-Area (the right side of Figure 2), and check black boxed behaviors recursively. Learners observe the behavior of a black boxed subordinate function, and check that these functions behave in the same way, which is one characteristic of recursive functions (this satisfies requirement for black boxing 2). When learners have noticed that the recursively called functions behave in the same way by repeating the disabling of black boxing, and want to stop the disabling, an instance of the calling function satisfying the termination condition is opened. Learners can then understand that there is a point in the execution process where recursive calling is terminated (this satisfies requirement for black boxing 3).

![Figure 2. 1-step Recursive Calling (left side) and Disabling Black Boxing (right side)](image)

2.5 Learning Scenario with TEDViT

We define an ideal learning scenario as steps I to IV described below. Steps I to III are for the learning in 2.4, while step IV is for the learning in 2.3; however, the system does not force learners to follow the scenario. It is preferable that the system is designed to let learners refer to messages to learn actively.
All recursively called functions are black boxed, and only first-executed functions are visualized during first step. Learners check an overview of the entire process without considering subordinate functions (Step I). Learners observe the recursive function characteristic that functions behave in the same way (Step II). Learners observe the termination of the recursive calling, in order to understand when the termination condition is satisfied (Step III). Learners check the passive flow by observing the processes of the functions that are black boxed in step I return outputs (Step IV).

3. Learning Flow Using the System

Learners learn recursive functions with C language. The system analyzes a source code and visualizes a program’s behavior. We extended TEDViT with the features in section 2. The interface of the system consists of four areas: the source code, message, control button, and STW areas (Figure 3). Learners can see the source code in the source code area. In the message area, learners can read messages suggesting what to do next or stating what type of processing was executed in the recursive calling. Learners can select which point in the execution process to be visualized by buttons in the control button area, and a process tree of recursive functions is produced in the STW area. Figure 4 (A) displays a situation where the first-called function calls functions recursively. The called function’s behaviors are black boxed with 1-step recursive calling (learning scenario Step I). Learners can observe black boxed recursive functions by clicking on their F-Areas. Figure 4 (B) shows a situation in which the recursive calling reaches the termination condition by repeatedly disabling black boxing (Step II). Then, a message suggests terminate condition and passive flow (Step III). Finally, learners check passive flow by observing the passing of control back from terminated functions (Step IV) (Figure 4 (C)).

Figure 3. System Interface

Figure 4. (A) 1-step Recursive Calling with Black Boxing, (B) Recursive Calling Reaching Termination Condition, (C) Observation of Passive Flow
4. Evaluation

4.1 Method of the Experiment

In order to evaluate whether the extended TEDViT is effective for learning recursive functions, we tested the three hypotheses below: The new TEDViT is effective for learning the execution process of a recursive function (Hypothesis 1). It is effective to visualize recursive calling and dataflow as a tree structure when TEDViT visualizes recursive calling behaviors (Hypothesis 2). Visualizing simplified behaviors of recursive calling by means of black boxing is effective for learning recursive calling (Hypothesis 3). We furthermore compared TEDViT with ANIMAL, because their methods are similar.

Subjects were eight students who have learned the C language and recursive functions. Subjects were divided into the control group and the experimental group based on the pretest result. Each group has four subjects. The control group learned recursive functions with ANIMAL, and the experimental group learned with the extended TEDViT. Subjects attended a brief review lecture on recursive functions at the beginning of the experiment, then took a pretest. Subjects in both groups learned with a learning support system, and afterwards they took a posttest. They learned recursive functions with the twice: Fibonacci sequence and merge sort. Finally, the subjects answered a questionnaire. The control group used ANIMAL, which includes visualization samples of both the Fibonacci sequence and merge sort. To confirm hypotheses 2 and 3, visualization with the tree structure and black box was included only when subjects learned the Fibonacci sequence.

We set the same questions for the pretest and posttest. The test consisted of three parts, each of which contained two questions, where the first part was about the Fibonacci sequence and the second about merge sort. The final part was to observe whether learners could apply the knowledge they gained to the system they used. The questions in the first and second parts were about programs with different parameters that the students observed with the systems, and we checked whether they successfully learned by using the systems according to the test results. In each part, the first question was to evaluate whether the students could explain recursive function behaviors with the concept of the black box, and the second was to evaluate whether they answered the return value and order of the return value being output (in short, it checked whether they understood passive flow or not).

Since the pretest and posttest contained the same questions, we could expect that all subjects would obtain higher scores in the posttest. We verified the learning effectiveness by checking the improvement ratio of the scores. The evaluation criteria for each question are as follows. Criteria for the first question: (1) Subjects can see recursive calling as a black box, (2) They can answer which value is returned from the black box, (3) They can explain how to calculate using the value from the black box and which value the function in question returns to the function that called it. Criteria for the second question is that whether they can illustrate the passive flow with returned values of each called function.

4.2 Results

Figure 5 illustrates the average pretest and posttest scores (maximum score is 100) and increase ratio. Experimental group got higher score at the posttest than the control group. The experimental group also marked higher increase ratio.

![Figure 5. Average Pretest and Posttest Scores (left side) and Increase Ratio (right side)]
Table 1 displays the questionnaire results. The control group’s scores for Q2 and Q3 are lower than those of the experimental group, which we believe resulted in the control group obtaining lower scores on the posttest. The control group’s score for Q4 is higher than that of the experimental group, which we consider being due to not only TEDViT, used by the experimental group, but also ANIMAL, used by the control group, including the feature of visualization with a tree structure.

Table 1: Questionnaire results

<table>
<thead>
<tr>
<th>Contents of questionnaire (best score is 5.00pt)</th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: It was easier to understand recursive functions than by learning only with a classroom lecture.</td>
<td><strong>4.25</strong></td>
<td>3.75</td>
</tr>
<tr>
<td>Q2: You could understand the idea of the black box.</td>
<td><strong>3.88</strong></td>
<td>2.88</td>
</tr>
<tr>
<td>Q3: You could understand passive flow.</td>
<td><strong>4.50</strong></td>
<td>3.13</td>
</tr>
<tr>
<td>Q4: You thought visualization of the recursive function execution process was understandable.</td>
<td>4.00</td>
<td><strong>4.50</strong></td>
</tr>
</tbody>
</table>

The results indicate that the extended TEDViT can provide effectiveness for learning the recursive function execution process, because the experimental group obtained higher scores and a more increased score rate than the control group. Furthermore, it is suggested that visualizing the recursive function execution process with a tree structure is effective, and it is necessary to grasp the idea of the black box and passive flow for understanding recursive functions.

5. Conclusion

We expanded TEDViT in order to achieve learning of recursive functions with the flexible visualization TEDViT originally applied. The results of our evaluation suggest that the proposed method is effective; however, potential points of improvement for the learning scenario were also revealed. Furthermore, more subjects are required to obtain more convincing results. Moreover, the current TEDViT is still not capable of managing structures. Since recursive programs include aspects such as tree structures, we are planning to expand TEDViT further to be able to manage structure.

Acknowledgements

This research was supported by the Japanese Grant-in-Aid for Scientific Research (B) Grant Number JP24300282, (C) Grant Number JP16K01084 and Grant-in-Aid for Young Scientists (B) Grant Number JP15K16104.

References


Can Distributed Practice Improve Students’ Efficacy in Learning their First Programming Language?

Qiujie Zhangb, Lishan Zhanga,b*, Baoping Lih,b, Ling Chena,b, I-Han Hsiaoc & Fati Wub
aBeijing Advanced Innovation Center for Future Education, Beijing Normal University, China
bFaculty of Education, Beijing Normal University, China
cSchool of Computing, Informatics & Decision System Engineering, Arizona State University, USA
*lishan@bnu.edu.cn

Abstract: Learning how to program has become required for many majors in higher education. However, programming is not easily learned, especially for non-engineering students. To improve students’ learning efficiency, we applied distributed practice to a C programming class with 69 college students in first grade, but have students decide the space of practice by themselves. By mining the relationships between practice patterns and learning performance, we found that students who practiced with high frequency significantly outperformed those who practiced infrequently. The frequency of practice was a strong predictor of both homework grades ($p=0.001$) and midterm exam grades ($p=0.023$). By contrast, the total amount of practice had very little effect on learning performance. The result shows that distributed practice is a better learning strategy than massed practice in C programming language learning. But the optimized space of practice in this domain has not been completely revealed yet.

Keywords: distributed practice, massed practice, programming language learning, higher education

1. Introduction

Mastering a programming language has become a required skill for many students in higher education, because programming now is an essential skill for data analysis in majors like economics, chemistry, biology and the social sciences. However, programming is not easily learned, especially for the students whose majors are non-engineering. These students usually only have programming related courses once a week, and do not have enough stimulus to practice. Therefore, if students can be encouraged to practice between the weekly classes, they may learn more efficiently.

Distributed practice is widely used in memorization (Dempster 1988; Rohrer 2015). For this task, the optimal space of learning or recalling is around one month. In the domain of language learning and reading comprehension, the optimal space tends to be reduced to one week or even shorter (Glenberg 1976; Balota, Duchek, & Paullin 1989; Begg & Green 1988). Previous studies seldom applied distributed practice in the domains like mathematics that were full of procedural knowledge (Cepeda, Pashler, Vul, Wixted, & Rohrer 2006), because space seems not always to have a significantly positive effect on learning procedural knowledge. Is distributed practice still useful in improving programming language learning efficiency? This paper studies whether distributed practice can help students learn C programming language, a typical domain full of procedural knowledge.

A previous study (Alzaid, Trivedi, & Hsiao, In press) introduced a system called “Quizit” which provided one multiple choice question for students to practice each day. Its preliminary result showed that studying actively with the system could lead to higher learning performance, but it was not clear whether it was due to the higher volume of practice, appropriate practice pattern or both. To further understand the effect of distributed practice on programming language learning, and also to test whether this bite-sized distributed practice could also help Chinese students, we adopted Quizit in
a C programming class which had 69 Chinese college students in their first grade. Moreover, we analyzed the relationship between students’ system usage pattern and their scores in midterm exams to answer the following research question: What factors in distributed practice can affect students’ learning performance? By knowing the answer to this research question, we can teach students how to use the system more efficiently in the future to improve their learning.

2. Experiment Design and Analysis Model

Quizit is a system that provides students a multiple choice question to practice every day. The user interface is very simple and straightforward. As soon as a student logs into the system, the question of the day will show up if the student has not finished it. Students can always retry or review the question of the day at their convenience. Students may not log into the system every day, so they do not always finish their questions on time. In this case, they can make up their unanswered questions by accessing the calendar view and navigating to the previous dates. All the multiple choice questions in the system are posted by the programming language instructor in advance. The system sends out the questions based on the order enforced by the instructor. A student can make comments on each question, and the comments can be seen by himself/herself immediately. But a student cannot see the comments made by others until he/she makes his/her own.

Every interaction made by students is recorded by the system, and each interaction is associated with a timestamp. The recorded interactions for the follow-up analysis include: correct attempt, incorrect attempt, question review (only look at the question and its correct answer), question retry (re-answer the question), comment on a question.

We conducted the experiment in a classroom scale. The class was C programming language learning, and had 69 college students in their first year, majoring in Psychology. According to power analysis (Faul, Erdfelder, Buchner, & Lang 2009), our sample size (N=69) is big enough (Power=0.849) to find out a correlation whose coefficient is 0.35 with two-tailed significance (p=0.05). The experiment lasted 41 days, which entailed that students were required to answer 41 questions in total. Although students had one question to practice per day, they attended the class only once a week. The class lasted 210 minutes every time, with 10 minutes break each 45 minutes. This was the very first class in computer programming for most of the students. Students were required to use the system along with class study, but the instructor did not enforce how often they should practice with the questions.

To mine the relationship between students’ system usage pattern and their learning performance, we designed nine indicators that represented different system usage behaviors and took the scores of four homework assignments as well as the midterm to quantify learning performance. With these indicators, we used Pearson correlation to find the relationships. The indicators are introduced rest of this section, and the analysis result is reported in the next section.

Because students can re-access their already answered questions, we calculated the measures to describe how they reviewed the questions. Specifically, two indicators were calculated for each student: the ratio of reviews, and the ratio of retries. Because retry is presumably more constructive than review, we expected that a higher number of retries may lead to better performance. We also explored students’ levels of system usage from two different dimensions: the aggregated system usages and system usage frequency. The total number of questions a student answered was used to approximate the aggregated system usages of the student. On the other hand, we used a counter to record how many days a student practiced with the system. Every time a student logged into the system and answered a question, the counter was increased by 1. The counter could only increase at most by 1 in a single day, no matter how many questions the student practiced. The interval of system usage for each student then could be calculated as the total number of days divided by the counter. The lower the interval of system usage was, the higher the system usage frequency was. In order to know in which situations students prefer to make comments, especially comments relevant to the question, we calculated five related indicators for each question, which included difficulty, total number of comments, number of meaningless comments, number of irrelevant comments, and number of relevant comments. Difficulty was calculated as [1 – percentage of correct attempts] with the range between 0 and 1, and it did not matter if the attempts were the first or not. If a question was never answered correctly by anyone, its difficulty was 1. If a question was always answered correctly, its
difficulty was 0. The other four indicators had their meaning labeled explicitly with their names. In addition, students’ performance using the system was quantified with the percentage of first attempt correctness and the overall percentage of correct attempts. The percentage of first attempt correctness represented a student’s base competence level regarding the related topic. By contrast, the overall percentage of correct attempts also took students’ performance in retrial into consideration.

3. Results

Besides the data collected from the system, we also had four homework grades and the midterm score for each student. These five scores were used to measure a student’s learning outcome. In this section, a descriptive result of the indicators is first reported, then the factors that affected learning performance are examined, and at last how a question’s features affected the comments elicited by it is explored.

3.1 Descriptive Results

In terms of treating already answered questions, most students like to review (93.87%) the questions, instead of retry (6.13%). Students did not log into the system to practice very often. On average, students used the system every 9.04 (SD=8.69) days. The mean value was heavily affected by 4 extreme students who never logged into the system at all. The medium interval of system usage was 6.83. This means most of the students practiced with the system at least once a week. The questions were not too easy or too difficult. The percentage of first check correctness was 0.601 (SD=0.165). The overall percentage of correctness was 0.691(SD=0.135). The 69 students only made 406 comments for all the 41 questions. Out of the 406 comments, there were 271 meaningless comments, 125 comments relevant to the questions, and 10 irrelevant comments. Students made many meaningless comments probably to check the comments made by others.

Table 1: The correlation between system usage indicators and learning outcomes (N=69)

<table>
<thead>
<tr>
<th>Outcome Indicator</th>
<th>Homework 1</th>
<th>Homework 2</th>
<th>Homework 3</th>
<th>Homework 4</th>
<th>Mean score of the 4 homework</th>
<th>Midterm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of reviews</td>
<td>r=-0.219</td>
<td>r=0.116</td>
<td>r=0.189</td>
<td>r=0.067</td>
<td>r=0.087</td>
<td>r=0.21</td>
</tr>
<tr>
<td></td>
<td>p=0.071</td>
<td>p=0.343</td>
<td>p=0.12</td>
<td>p=0.582</td>
<td>p=0.478</td>
<td>p=0.083</td>
</tr>
<tr>
<td>Ratio of retries</td>
<td>r=-0.025</td>
<td>r=-0.085</td>
<td>r=0.113</td>
<td>r=0.094</td>
<td>r=0.035</td>
<td>r=0.042</td>
</tr>
<tr>
<td></td>
<td>p=0.837</td>
<td>p=0.488</td>
<td>p=0.357</td>
<td>p=0.444</td>
<td>p=0.775</td>
<td>p=0.732</td>
</tr>
<tr>
<td>Total number of attempts</td>
<td>r=-0.182</td>
<td>r=-0.234</td>
<td>r=0.233</td>
<td>r=0.017</td>
<td>r=0.153</td>
<td>r=0.135</td>
</tr>
<tr>
<td></td>
<td>p=0.134</td>
<td>p=0.053</td>
<td>p=0.053</td>
<td>p=0.888</td>
<td>p=0.208</td>
<td>p=0.268</td>
</tr>
<tr>
<td>Interval of system usage</td>
<td>r=0.046</td>
<td>r=-0.205</td>
<td>r=-0.332**</td>
<td>r=-0.247*</td>
<td>r=-0.303*</td>
<td>r=-0.273*</td>
</tr>
<tr>
<td></td>
<td>p=0.71</td>
<td>p=0.091</td>
<td>p=0.005</td>
<td>p=0.041</td>
<td>p=0.011</td>
<td>p=0.023</td>
</tr>
<tr>
<td>Percentage of correct attempts</td>
<td>r=0.011</td>
<td>r=0.011</td>
<td>r=0.071</td>
<td>r=0.056</td>
<td>r=-0.004</td>
<td>r=0.109</td>
</tr>
<tr>
<td></td>
<td>p=0.927</td>
<td>p=0.931</td>
<td>p=0.564</td>
<td>p=0.648</td>
<td>p=0.972</td>
<td>p=0.372</td>
</tr>
<tr>
<td>Percentage of first attempts being correct</td>
<td>r=0.078</td>
<td>r=0.003</td>
<td>r=0.035</td>
<td>r=0.149</td>
<td>r=0.090</td>
<td>r=0.156</td>
</tr>
<tr>
<td></td>
<td>p=0.522</td>
<td>p=0.98</td>
<td>p=0.778</td>
<td>p=0.223</td>
<td>p=0.461</td>
<td>p=0.201</td>
</tr>
<tr>
<td>Ratio of meaningless comments</td>
<td>r=-0.012</td>
<td>r=-0.074</td>
<td>r=-0.186</td>
<td>r=-0.142</td>
<td>r=-0.165</td>
<td>r=0.133</td>
</tr>
<tr>
<td></td>
<td>p=0.92</td>
<td>p=0.544</td>
<td>p=0.126</td>
<td>p=0.244</td>
<td>p=0.176</td>
<td>p=0.275</td>
</tr>
<tr>
<td>Ratio of irrelevant comments</td>
<td>r=0.052</td>
<td>r=-0.125</td>
<td>r=0.065</td>
<td>r=0.04</td>
<td>r=0.003</td>
<td>r=-0.104</td>
</tr>
<tr>
<td></td>
<td>p=0.673</td>
<td>p=0.306</td>
<td>p=0.594</td>
<td>p=0.742</td>
<td>p=0.979</td>
<td>p=0.397</td>
</tr>
<tr>
<td>Ratio of relevant comments</td>
<td>r=-0.004</td>
<td>r=-0.138</td>
<td>r=-0.104</td>
<td>r=-0.113</td>
<td>r=0.021</td>
<td>r=0.18</td>
</tr>
<tr>
<td></td>
<td>p=0.974</td>
<td>p=0.259</td>
<td>p=0.397</td>
<td>p=0.357</td>
<td>p=0.866</td>
<td>p=0.14</td>
</tr>
</tbody>
</table>
3.2 Which Factors Affected Learning?

A student’s learning performance was evaluated from two aspects: homework grades and midterm scores. One student had four homework grades before taking the midterm exam. The average score of the four homework grades for each student was calculated. It ended up with six performance measures in total. Out of the six measures, a mean score of the four homework grades and the midterm score were considered to be the two most important measures. Note that students could help each other while doing homework but they had to work on their own while taking the midterm exam.

We calculated the Pearson correlation coefficient between the indicators and the six outcome scores respectively. The result is described in Table 1. It showed that all but one indicator had no correlation with the performance outcome. Interval of system usage was the only exception. It did not correlate to the grades of the first two homework assignments, but strongly correlated to all the other performance outcome measures. Considering that early homework just had students review basic concepts, the interval of system usage was believed to be an important factor affecting learning performance. Neither the way to re-access the answered questions nor distribution of different types of comments had an effect on students’ learning performance.

It is also interesting to see how the two different dimensions (i.e. close-book & close-discussion vs. open-book & open-discussion) of learning performance measures are correlated. The result of Pearson correlation is described in Table 2. Surprisingly, the scores of Homework 4 was the only indicator that was significantly correlated to midterm scores.

Table 2: The correlation between homework scores and midterm scores (N=69)

<table>
<thead>
<tr>
<th></th>
<th>Homework 1</th>
<th>Homework 2</th>
<th>Homework 3</th>
<th>Homework 4</th>
<th>Mean score of the 4 homework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm scores</td>
<td>r=0.147</td>
<td>r=0.071</td>
<td>r=0.117</td>
<td>r=0.321**</td>
<td>r=0.235</td>
</tr>
<tr>
<td></td>
<td>p=0.227</td>
<td>p=0.56</td>
<td>p=0.337</td>
<td>p=0.007</td>
<td>p=0.052</td>
</tr>
</tbody>
</table>

3.3 When do Students Want to Make Comments?

Pearson correlation coefficient was calculated between question difficulty and the number of their three different types of corresponding comments respectively. The correlation between question difficulty and the total number of corresponding comments was also calculated. The result is summarized in Table 3. The number of meaningless comments and total number of comments significantly correlated to question difficulty. At the same time, students were also actively making significantly more relevant comments to ask for help or share their thoughts. It appeared that students tended to refer to the comments for help when they had difficulty in answering questions correctly.

Table 3: The correlation between question difficulty and different types of comments (N=41)

<table>
<thead>
<tr>
<th></th>
<th>Number of meaningless comments</th>
<th>Number of relevant comments</th>
<th>Number of irrelevant comments</th>
<th>Total number of comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question difficulty</td>
<td>r=0.401**</td>
<td>p=0.009</td>
<td>r=0.355*</td>
<td>r=0.107</td>
</tr>
<tr>
<td></td>
<td>p=0.09</td>
<td></td>
<td>p=0.023</td>
<td>p=0.504</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r=0.467**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=0.002</td>
</tr>
</tbody>
</table>

4. Discussion

The result reveals several interesting findings. In terms of the learning factors, we found that neither number of practices, which was represented by total number of attempts, nor percentage of correct attempts was significantly related to learning outcomes. This seems inconsistent with the well-known Learning Factors Analysis (LFA) theory (Cen, Koedinger, & Junker 2006), which quantifies learning outcome with the number of correct attempts and incorrect attempts, and claims that learning outcomes should be improved with practice. However, different from the traditional setting of LFA, which provided more than enough practice questions, students in our experiment had a fixed number of practices.
(41) of questions and all the students answered the 41 questions at least once due to the requirement from the instructor. Students could practice with the same question as many times as they wanted, which led to the difference in the number of attempts. This experiment setting provides us an opportunity to see how to better use the fixed size of practice questions for learning. From the result, it did not matter how a student reviewed the questions (just look at it or do it again). This was probably because all the questions were multiple choice questions. We suspected that type of review might play a more important role for the questions that required student to do actual programming. As mentioned above, the number of practices also did not matter. The only significant factor was the interval of practice. The lower the interval was, the better the learning performance was. This again seems inconsistent with the previous conclusions about spacing effect and distributed practice (Cepeda et al. 2009; Cepeda, Vul, Rohrer, Wixted, & Pashler 2008). But it is not the case. Traditional spacing effect experiments fixed the number of opportunities that one could practice and gave participants the same amount of practice every time. But we provided the fixed size of practice resources for each student, and let students decide the distribution of their practice. If a student’s interval of practice was one day, he/she practiced one question per day, which was actually adding one-day space between each two questions. If a student’s interval of practice was seven days, he/she would practice seven questions at a time. So the student added no space between the seven questions, but a seven -day space between every seven questions. The result was consistent with the finding in comparing massed practice to distributed practice (Dempster 1988; Hopkins, Lyle, Hieb, & Ralston 2016).

We had face-to-face interviews with six students after the midterm exam to understand what prevented some students from doing exercises in the system frequently, and why distributed practice helped some of them. The six students covered four different types: frequent system usage with high midterm score, infrequent system usage with high midterm score, frequent system usage with low midterm score, and infrequent system usage with low midterm score. All the students claimed that time was the biggest issue that prevented them from practicing the exercises every day. They were busy with many other classes. If they did have time, all but one student would like to practice every other day. The lowest-performing student simply felt bothered logging into the system and answering questions. It was not a surprise that the student got a very low score on the midterm. The reason why students did not want to practice every day was due to the cost of mental context switching (Kieffäber & Hetrick 2005). They often needed to prepare themselves for answering the questions, especially for the questions containing programs. They said that it usually just took three to five minutes to answer one question, so it was not worthwhile to answer just one question after the students spent an equally long time to prepare themselves. On the other hand, because answering one or two questions should only take less than 10 minutes, students should be able to spare this amount of time if they are somehow encouraged to do so. How to encourage students is one of our next projects. We could not explain well why some students still got a low grade on the midterm although they practiced frequently. It could be because practicing exercises in the system was the only thing they did for studying in this class. Indeed, there were many other variables that affected their learning outcome that we could not capture. Also, the experiment was just enough to claim that relatively shorter interval of practice could lead to better learning outcomes in C programming language learning. But we cannot conclude the optimized space of distributed practice for this course work.

We were expecting students who made more relevant comments would achieve higher learning performance, but this was not the case. It was probably because the comment function was not well introduced and recognized by students. After all, there were only 125 relevant comments in total. Students used this function only when they met difficulties. As the class goes on, questions will become more and more challenging, so we are expecting to see more comments in the future. For initiating students’ discussions around the questions, we also considered compelling students to make comments.

Students spent less time answering simple questions and also did not have enough stimulus to post comments. By contrast, student had to spend a longer time dealing with complex questions, including answering and posting comments. This finding provides us an adaptive question recommendation mechanism. Based on an individual’s time schedule, the system could choose to post a simple question for a busy day, and post a complex question for a leisure day. The utility of a question can be calculated in terms of time schedule and question difficulty, and the utility value can be used for the question recommendation (Zhang & Vanlehn 2016).
5. Conclusion

By having students decide the space of practice by themselves, we found that when all the students had the same amount of questions to practice, the students who practice frequently significantly outperformed those who practice infrequently in both homework and midterm exam. Neither the number of times re-practicing nor the way of re-visiting the answered questions affected students’ learning performance. Many students claimed that they were short of time in practicing frequently, but the fact was that solving one simple multiple choice question usually only took less than five minutes. Complex questions usually need longer time to be solved, and students also tend to make more comments about these questions. So we should not expect students to spend equal amounts of time on all of the questions. In future work, we will consider how to encourage students and provide adaptive question recommendations based on students’ time schedules.

Acknowledgments

This work was supported by The Chinese Ministry of Education Humanities and Social Science Foundation under projects number 14YJC880025, and China Postdoctoral Science Foundation under projects number 2017M610054.

References


Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. American Psychologist, 43(43), 627-634


432
Cross Analytics of Student and Course Activities from e-Book Operation Logs

Atsushi SHIMADA* & Shin’ichi KONOMI
*Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
bFaculty of Arts and Science, Kyushu University, Japan
*atsushi@ait.kyushu-u.ac.jp

Abstract: In this paper, we propose a cross analytics methodology of student activities and course activities using e-Book operation logs collected in 15 courses with face-to-face lecture style over 4 weeks. These courses commonly use the same lecture materials, but are conducted by different teachers. The new aspect of our research is that we perform cross analysis over courses. Most past researches focus on students’ activities in a specific course, and give discussions about how the students behaved, how the behaviors differ from each other. In contrast, our research focuses on the course activities and conducts a comparison among courses. First, we begin with data alignment for row data to rectify a student activity every 10 seconds. Through our analytics, it becomes clear that whether students’ activities varies with teachers or their teaching styles. In the experiments, we applied the proposed analytics to 1.1-million operation logs, and found out interesting characteristics through the comparison across courses.

Keywords: Learning analytics, educational data mining, e-Book logs, learning activity

1. Introduction

Much attention has been paid to learning analytics, which is defined as the measurement, collection, analysis, and reporting of data about learners and their context for the purpose of understanding and optimizing learning and environments in which it occurs (https://solaresearch.org/). Thanks to digital learning systems such as Blackboard and Moodle, large scale educational data can be easily collected. This has changed the trend of research from data measurement and collection to analytics and feedback for improvement of teaching and learning.

One of the major studies of educational data analytics is to understand students’ activities during a course period. Many studies have focused on clickstream data collected from Massive Open Online Courses (MOOCs), and analyzed the data for prediction of course completion (Crossley 2016), understanding students’ learning path (Davis 2016), discovering relationship between activities and culture (Liu 2016), and change detection of students’ behavior (Park 2017). Another possible educational data is available on an e-Book system (a digital textbook system) and e-Learning system. For example, the data was analyzed for pattern mining of preview and review activities (Oi 2015), understanding learning behavior of students (Yin 2015), browsing pattern mining (Shimada 2016), and performance prediction (Okubo 2016). As introduced above, learning analytics has been addressed from various aspects, but mainly has been focused on a specific course.

In this paper, we focus on e-Book operation log data collected from approximately 2,700 students across 15 courses over 4 weeks. There are approximately 1.1 million operation logs that record students’ time-series activities during lecture periods. It must be noted that all the courses follow the same course design, but are conducted by different teachers. Therefore, it is possible to analyze students’ activities not only within a specific course, but also across courses. The aim of this study is to investigate whether students’ activities varies with teachers or their teaching styles. First, we apply a data alignment technique that can cleanse data for the ease of analytics. Then, we analyze the browsing speed of each individual student. Following this, we apply a simple anomaly detection method to the time-series operation logs to discover changes over time for each student. We summarize the individual and course activities during a lecture in the form of an 8-dimensional activity vector containing e-Book operation features, abnormal states, and browsing speeds.
2. Datasets of e-Book Operation Logs

The dataset that we used in our study was collected from the Kyushu University e-Learning system and e-Book system. The target course is a series of lectures that constitute the “Primary Course of Cyber Security,” which commenced in April 2017. All first-year students (more than 2,600 students, including both arts and science students) in Kyushu university are required to take this course. All students have their own laptops and bring them to access the e-Learning system and e-Book system during the lecture every week. There are a total of 15 courses (approximately 180 students/course on an average), in which teachers use the same lecture materials.

Table 1 shows the detailed course information: course id, teacher, and number of students. Note that five teachers are assigned two lectures each (e.g., Te01 is assigned C01 and C04). We analyzed e-Book operation logs collected over a 4-week period (approximately 1 million operation logs). For courses C08 and C13, which were conducted by Te06, the number of operation logs in the 3rd and 4th week eventually decreased. According to the teacher, the reason is that the e-Book system did not work well in the 3rd week. Thus, the teacher gave students a PDF version of the lecture material. In the 4th week, the teacher presented a different topic from that prescribed for the other courses. We ignored the logs of these two courses in the 3rd week and 4th week in our analysis.

Table 1: The details of course information and the number of e-Book logs in each course

<table>
<thead>
<tr>
<th>Course</th>
<th>Teacher</th>
<th>Students</th>
<th>1st week</th>
<th>2nd week</th>
<th>3rd week</th>
<th>4th week</th>
<th>Course Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>C01</td>
<td>Te01</td>
<td>208</td>
<td>14,186</td>
<td>12,927</td>
<td>13,131</td>
<td>10,495</td>
<td>50,739</td>
</tr>
<tr>
<td>C02</td>
<td>Te02</td>
<td>185</td>
<td>11,839</td>
<td>42,130</td>
<td>33,777</td>
<td>28,859</td>
<td>116,605</td>
</tr>
<tr>
<td>C03</td>
<td>Te03</td>
<td>208</td>
<td>23,267</td>
<td>30,820</td>
<td>50,910</td>
<td>48,107</td>
<td>153,104</td>
</tr>
<tr>
<td>C04</td>
<td>Te01</td>
<td>157</td>
<td>9,092</td>
<td>28,301</td>
<td>22,082</td>
<td>15,985</td>
<td>75,460</td>
</tr>
<tr>
<td>C05</td>
<td>Te04</td>
<td>192</td>
<td>18,862</td>
<td>9,185</td>
<td>18,460</td>
<td>13,571</td>
<td>60,078</td>
</tr>
<tr>
<td>C06</td>
<td>Te05</td>
<td>175</td>
<td>7,423</td>
<td>4,418</td>
<td>14,777</td>
<td>12,506</td>
<td>39,124</td>
</tr>
<tr>
<td>C07</td>
<td>Te03</td>
<td>175</td>
<td>16,029</td>
<td>18,702</td>
<td>42,674</td>
<td>39,029</td>
<td>116,434</td>
</tr>
<tr>
<td>C08</td>
<td>Te06</td>
<td>200</td>
<td>18,983</td>
<td>22,737</td>
<td>1,541</td>
<td>235</td>
<td>43,496</td>
</tr>
<tr>
<td>C09</td>
<td>Te02</td>
<td>137</td>
<td>2,473</td>
<td>20,343</td>
<td>20,689</td>
<td>15,167</td>
<td>58,672</td>
</tr>
<tr>
<td>C10</td>
<td>Te07</td>
<td>176</td>
<td>21,195</td>
<td>19,361</td>
<td>24,215</td>
<td>23,721</td>
<td>88,492</td>
</tr>
<tr>
<td>C11</td>
<td>Te08</td>
<td>144</td>
<td>33,890</td>
<td>22,698</td>
<td>27,486</td>
<td>25,488</td>
<td>109,562</td>
</tr>
<tr>
<td>C12</td>
<td>Te05</td>
<td>180</td>
<td>9,246</td>
<td>3,825</td>
<td>10,848</td>
<td>8,498</td>
<td>32,417</td>
</tr>
<tr>
<td>C13</td>
<td>Te06</td>
<td>178</td>
<td>22,564</td>
<td>19,103</td>
<td>304</td>
<td>18</td>
<td>41,989</td>
</tr>
<tr>
<td>C14</td>
<td>Te09</td>
<td>213</td>
<td>20,002</td>
<td>6,568</td>
<td>3,808</td>
<td>17,319</td>
<td>47,697</td>
</tr>
<tr>
<td>C15</td>
<td>Te10</td>
<td>171</td>
<td>24,482</td>
<td>30,891</td>
<td>26,968</td>
<td>20,773</td>
<td>103,114</td>
</tr>
<tr>
<td><strong>Week Subtotal</strong></td>
<td></td>
<td><strong>253,533</strong></td>
<td><strong>292,009</strong></td>
<td><strong>311,670</strong></td>
<td><strong>279,771</strong></td>
<td><strong>1,136,983</strong></td>
<td></td>
</tr>
</tbody>
</table>

3. Proposed Method

3.1 Data Alignment

When an e-Book is operated, its timestamp, user id, material id, page number and operation name are automatically recorded as an operation event. There are many types of operations; for example, OPEN indicates that a student has opened the e-Book file and NEXT indicates that the student has clicked the next button to move to the subsequent page. Students can bookmark a specific page, highlight selected characters, and make notes on a page. These operations correspond to the events ADD BOOKMARK, ADD MARKER, and ADD MEMO, respectively. There are two issues to consider when data analysis is performed for these event logs. First, the operation timing of students are not completely synchronized. We have to consider a slight time delay of operation. Second, the e-Book system only records the event logs and does not know a student’s activities between successive event logs. We have to complement the missing activities.
Let $N$ be the number of students in a course, and $i$ be an index that refers to an individual student such that $i = 1, ..., N$. We assume discrete time, where each time-point is associated with a short period called time window. In other words, a 90-min time period is divided into short time windows, where $t = 1, ..., T$ is an index that runs from the first to the last time period. In our study, we define the length of time window to be 10 seconds; therefore, $T$ equals 540 (because 90 minutes = 5,400 seconds, $5,400/10$ seconds = 540 time windows).

Let $S_{i,t}$ be the state of student $i$ at $t$. $S_{i,t}$ consists of five status: $S_{i,t} = (e_{i,t}, p_{i,t}, b_{i,t}, m_{i,t}, n_{i,t})$, where each element stores the following information.

- $e_{i,t}$: the number of events operated by student $i$ at $t$
- $p_{i,t}$: the page number browsed by student $i$ at $t$
- $b_{i,t}$: the number of events of type “ADD BOOKMARK” operated by student $i$ at $t$
- $m_{i,t}$: the number of events of type “ADD_MARKER” operated by student $i$ at $t$
- $n_{i,t}$: the number of events of type “ADD_MEMO” operated by student $i$ at $t$

After the alignment, the status of every student can be represented by the uniformly sized vector $S_{i,t}$.

### 3.2 Pace Estimation

It is meaningful to know how many students open and browse the same page, and how many students are delayed. We assume that most students are following teacher’s explanation and reading the same page at $t$, so $M_t = \text{median}(p_{i,t})$ is calculated as the page taken to be the one being read by the majority of all students. Let $v_{i,t}$ be the pace of student $i$ at $t$, whose value is calculated by the following formula.

$$v_{i,t} = p_{i,t} - M_t$$

If $v_{i,t}$ is a negative value, it means that student $i$ is behind; otherwise, the student is ahead compared with the majority of students taking the course. Figure 1 depicts a visualization of students’ paces in a 90-min lecture. The blue color and the red color correspond to $v_{i,t} < 0$ and $v_{i,t} > 0$, respectively. When $v_{i,t} = 0$, the painted color is black. In the ideal case, all students open the same page at the time, and the visualization result is a black colored rectangle. In Figure 1, most of students browsed the same page at the beginning part of the lecture. From the middle part, some students began to be delayed. In our experiments below, we compared the speed status among courses across 4 weeks.

![Figure 1. Visualization of browsing/reading speed in 90min lecture](image)

### 3.3 Anomaly Detection

Change detection or anomaly detection is a useful technique used to discover a student who is doing something else when compared with other students. In our study, we detect two kinds of anormal states of a student. One is the anormal page view, which indicates that a student is browsing a different page...
from the majority. To determine whether the page view is anormal or not, we introduce a threshold $t_{th}$.

Let $mpv_{i,t}$ be the anormal page view state of student $i$ at $t$, whose value is defined as follows.

$$mpv_{i,t} = \begin{cases} 1, & |v_{i,t}| > t_{th} \\ 0, & |v_{i,t}| \leq t_{th} \end{cases}$$

The other is an anormal event number that indicates the state of a student with less operation logs than those of other students. Let $men_{i,t}$ be the anormal event state of student $i$ at $t$.

$$men_{i,t} = \begin{cases} 1, & e_{i,t} < me_{t} - se_{t} \\ 0, & otherwise \end{cases}$$

where $me_{t}$ and $ms_{t}$ represent the mean value and standard deviation of $e_{i,t}$ of all the students in a course. The anomaly detection results are aggregated to summarize a student’s activity within a 90-min lecture. The details are explained in the following section.

### 3.4 Student Activity and Course Activity

The individual student activity $a_{i}$ is defined as

$$a_{i} = \left( \frac{\sum_{t} b_{i,t}}{\sum_{t} e_{i,t}}, \frac{\sum_{t} m_{i,t}}{\sum_{t} e_{i,t}}, \frac{\sum_{t} n_{i,t}}{\sum_{t} e_{i,t}}, \frac{\sum_{t} apv_{i,t}}{T}, \frac{\sum_{t} aen_{i,t}}{T}, \frac{\sum_{t} f(slower)}{T}, \frac{\sum_{t} f(normal)}{T}, \frac{\sum_{t} f(faster)}{T} \right),$$

where $f(X)$ is a function that computes the number of time windows that satisfy a given condition $X$. The three conditions are defined as follows.

- $slower = v_{i,t} < -t_{th}$
- $normal = |v_{i,t}| \leq t_{th}$
- $faster = v_{i,t} > t_{th}$

The activity $a_{i}$ consists of eight elements. The first three elements are ratios of three operations related to bookmark, marker, and memo. The fourth and fifth elements are the ratio of the number of time windows detected as anormal states over time. The last three elements are the ratios of time windows over time when above three speed conditions are satisfied.

The course activity $C$ is acquired by averaging $a_{i}$ in the course. $A_{i}$ can be written as

$$C = \frac{\sum_{t} a_{i}}{N},$$

where the numerator calculates elementwise summation. The course activity can be used as a barometer of a specific course, and can be easily compared with those of other courses. In our study, we calculated 15 course activities over 4 weeks.

### 4. Experimental Results

We applied our analytics method to the rectified data (alignment data in 540 time windows) of all courses. Figure 2 shows the visualization results of speed analytics for the 4 week datasets. The red/blue color denotes that the reading/browsing speed of the student was faster/slower than the majority of the students. The brighter the color, the larger the degree of pace.

Each week, the color of speed visualizer is different among the courses. For instance, the main color of C04 is red, while C07 has a predominantly blue color. We categorized the speed characteristics into the following four types. A description is provided for each.

Type1: Most students were browsing a page at the same pace with other students. The visualized result is shown in black. For example, C08, C10, and C13 in the 1st week.

Type2: Many students were delayed in browsing. The visualized result is shown in blue color for a 90-min period. For example, C06, C12, and C14 in the 3rd week.

Type3: Many students browsed later pages as the lecture went on. The visualized color changes from blue (or dark blue) to red. For example, C02, C03, and C11 in the 1st week.

Type4: Many students were getting delayed gradually as the lecture went on. The visualized color changes from red (or dark red) to blue. For example, C04 and C13 in the 2nd week.
As summarized above, the characteristics of courses are obviously different from each other despite the fact that all of them use the same lecture organization and materials. We suppose that the teacher’s personality influences the progression of a lecture. Five teachers were assigned two courses each; C01 and C04 by Te01, C02 and C09 by Te02, C03 and C07 by Te03, C06 and C12 by Te05, and C08 and C13 by Te06. Among these, courses conducted by the same teacher were found to have similar characteristics. For example, the color maps of C03 and C07 are very similar when compared with those of other courses. Across the weeks, the similar color maps were drawn in each row, which implied that the method of conducting a lecture strongly depended on the teacher.

In Table 2, a summary of eight kinds of activities (i.e., course activity C explained in Section 3.4 for each course) in the 1st week and the average of 4 weeks is shown. In the figure, a larger score in each column is highlighted using red color. The bookmark function was only used by the students in C1, C02, C03, C04, and C07 in the 1st week. The memo function was also used in limited courses. On the other hand, the marker function was used by students in all courses. This might be because the marker function is easy to use when compared with the other two functions. A student only has to trace a cursor over characters after selecting the marker function. Although the bookmark function does not require a complicated operation, a marker function is convenient in terms of marking specific sections (not a page but a content in the page). In contrast, when a student uses a memo function, it will take a

![Figure 2. Comparison of speed status across courses and weeks](image)

<table>
<thead>
<tr>
<th>Course</th>
<th>Teacher</th>
<th>1st week</th>
<th>2nd week</th>
<th>3rd week</th>
<th>4th week</th>
</tr>
</thead>
<tbody>
<tr>
<td>C01</td>
<td>Te01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.25</td>
<td>0.62</td>
</tr>
<tr>
<td>C02</td>
<td>Te02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.13</td>
<td>0.62</td>
</tr>
<tr>
<td>C03</td>
<td>Te03</td>
<td>0.00</td>
<td>0.03</td>
<td>0.31</td>
<td>0.38</td>
</tr>
<tr>
<td>C04</td>
<td>Te04</td>
<td>0.00</td>
<td>0.01</td>
<td>0.38</td>
<td>0.75</td>
</tr>
<tr>
<td>C05</td>
<td>Te05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C06</td>
<td>Te06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.38</td>
</tr>
<tr>
<td>C07</td>
<td>Te07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C08</td>
<td>Te08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.38</td>
</tr>
<tr>
<td>C09</td>
<td>Te09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C10</td>
<td>Te10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C11</td>
<td>Te11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C12</td>
<td>Te12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C13</td>
<td>Te13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C14</td>
<td>Te14</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>C15</td>
<td>Te15</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.32</td>
</tr>
</tbody>
</table>

| Table 2: Summary of activities in each course. Left: 1st week, Right: average of 4 weeks |}

![Table 2](image)
longer time to type characters on the form. Therefore, a student has to complete typing as quickly as possible. Otherwise, he/she will be delayed in comparison to the pace of lecture.

The course activities throughout the 4 weeks exhibited tendencies similar to that of the activities in the 1st week. In fact, the correlation of each activity between the 1st week and average of 4 weeks was very high (0.92, 0.72, 0.85, 0.66, 0.73, 0.67, 0.65, 0.41 for each of the activities). The findings suggest that a specific course activity does not change much over the weeks, which implies that each teacher has his/her own teaching style.

5. Conclusion

In this paper, we analyzed e-Book operation logs collected from approximately 2,700 students enrolled in 15 courses over 4 weeks. We developed new analytics methodologies and applied them to the 1.1 million log data entries. The new findings from this study are summarized below. The approach to teaching differed even though teachers followed the same course design and used the same lecture materials. Each teacher tended to keep his/her own teaching style over weeks.

In future work, we plan to analyze the analytics of e-Book operation logs across other courses over years to establish the ease of use of the proposed method for long-term and wide-spread educational data. In addition, we will develop a feedback system for teachers and students. A feedback regarding real-time anomaly detection will be useful for teachers to change the pace or provide detailed explanation during lectures. It will be helpful for students to know the activities of other students and compare self-activity with other students. Through feedback using quantitatively represented learning activities, we would like to improve education in the big data era.

References

Students’ Performance Prediction Using Data of Multiple Courses by Recurrent Neural Network

Fumiya OKUBO*\textsuperscript{a}, Takayoshi YAMASHITA\textsuperscript{b}, Atsushi SHIMADA\textsuperscript{c} & Shin’ichi KONOMI\textsuperscript{a}

\textsuperscript{a} Faculty of Arts and Science, Kyushu University, Japan  
\textsuperscript{b} Department of Computer Science, Chubu University, Japan  
\textsuperscript{c} Faculty of Information Science and Electrical Engineering, Kyushu University, Japan  
* fokubo@artsci.kyushu-u.ac.jp

Abstract: In this paper, we show a method to predict students’ final grades using a recurrent neural network (RNN). An RNN is a variant of a neural network that handles time series data. For this purpose, the learning logs from 937 students who attended one of six courses by two teachers were collected. Nine kinds of learning logs are selected as the input of the RNN. We examine the prediction of final grades, where the training data and test data are the logs of courses conducted in 2015 and in 2016, respectively. We also show a way to identify the important learning activities for obtaining a specific final grade by observing the values of weight of the trained RNN.

Keywords: learning analytics, recurrent neural network, learning log, prediction of student’s performance

1. Introduction

At Kyushu University, a learning support system, called M2B system, was introduced in 2014, and a mechanism for recording various logs of learning activities was developed. All students in Kyushu University have their own PC and they can use M2B system both inside and outside classroom. Hence, a large amount of their learning logs are accumulated in M2B system. By March 2017, about 45 million logs have been collected in M2B system. Using these learning logs, a number of investigations on learning analytics have been conducted at Kyushu University (e.g., Ogata et al. (2015), Ogata et al. (2017), Okubo et al. (2016), Okubo et al. (2015)).

In the field of educational data mining, predicting students’ performance is important topic, which is useful for teachers and students. Particularly, an early detection of students who are likely to fail or drop out of class, i.e., students referred to as “at-risk” students has been intensively investigated such as in Hlosta, Zdrahal & Zendulka. (2017). In Marbouti, Diep-Duc & Madhavan. (2016), the prediction method for identifying “at-risk” students in educational data mining is summarized, where Logistic Regression, Support Vector Machine, Decision Tree, Multi-Layer Perceptron, Naïve Bayes Classifier, and K-Nearest Neighbor are guided.

We consider the prediction of students’ performance using an artificial neural network which has recently become a hot topic. With a general neural network, it is impossible to consider the relationship between past data and current data in the time series. This situation often appears in the case of learning logs. A recurrent neural network (RNN) is a variant of a neural network, that can address this problem (Bodén, M. (2002)). An RNN can obtain the output value based on past and current information using the internal loops of the network. In Okubo et al. (2017), we proposed a method for predicting students’ performance with an RNN from the log data collected in M2B system. It was confirmed that an RNN is effective in the early prediction of final grade of students in cases where training data and test data were collected from the same course. In order to apply the results of prediction at actual education sites, the model must be trained using data on courses that were held in the past. In addition, the teacher of the course for the training data may be different from the one who will consult the course to be predicted.
In this paper, for the sake of validation of the method of the prediction by an RNN in a more realistic situation, we collect learning logs from six courses. These courses were conducted by two different teachers in 2015 and 2016, who followed the same syllabus for eight week. We examined the method of prediction by an RNN by using the data on courses held in 2015 and 2016 as the training data and the test data. We also compare the results with predictions by multiple regression analysis.

2. Data Collection

2.1 Active Learner Point

Many kinds of logs of learning activities are stored in M2B system. For analyzing and visualizing these data easily, we select nine major learning activities, and evaluate them of each student from 0 to 5 points for each week of a course. A vector of these nine evaluations is called an Active Learner Point (ALP). The nine selected learning activities are summarized in Table 1. We note that

- The logs of a total time of slide views are only calculated for viewing outside of class.
- The logs of the number of course views, total time of slide views, number of markers, number of actions, and word count in a journal are evaluated relatively in a course. For example, the top 10% of students who are doing a lot about the learning activity obtain a score of 5.

<table>
<thead>
<tr>
<th>Activities</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>Attendance</td>
<td>Being late</td>
<td></td>
<td></td>
<td></td>
<td>abs.ence</td>
</tr>
<tr>
<td>Quiz</td>
<td>Above 80%</td>
<td>Above 60%</td>
<td>Above 40%</td>
<td>Above 20%</td>
<td>Above 10%</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Report</td>
<td>Submission</td>
<td>Late</td>
<td></td>
<td></td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>Course views</td>
<td>Upper 10%</td>
<td>Upper 20%</td>
<td>Upper 30%</td>
<td>Upper 40%</td>
<td>Upper 50%</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Slide views</td>
<td>Upper 10%</td>
<td>Upper 20%</td>
<td>Upper 30%</td>
<td>Upper 40%</td>
<td>Upper 50%</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Markers</td>
<td>Upper 10%</td>
<td>Upper 20%</td>
<td>Upper 30%</td>
<td>Upper 40%</td>
<td>Upper 50%</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Memos</td>
<td>Upper 10%</td>
<td>Upper 20%</td>
<td>Upper 30%</td>
<td>Upper 40%</td>
<td>Upper 50%</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Actions</td>
<td>Upper 10%</td>
<td>Upper 20%</td>
<td>Upper 30%</td>
<td>Upper 40%</td>
<td>Upper 50%</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Word count</td>
<td>Upper 10%</td>
<td>Upper 20%</td>
<td>Upper 30%</td>
<td>Upper 40%</td>
<td>Upper 50%</td>
<td>Otherwise</td>
</tr>
</tbody>
</table>

2.2 Courses

We collected learning logs regarding ALPs from six “Information Science” courses for first grade students, as shown in Table 2. These courses were conducted, basically, by following the same syllabus for eight weeks, i.e., they are quarter courses. These courses were conducted in different years and terms with different teachers, and participating students. The prepared materials were needed to implement a standard plan for “Information Science” courses. The teachers were permitted to change this plan so long as the objectives of the course were not changed, e.g., by changing the order of subjects to be taught, adding the extra teaching materials, reports, quizzes, and so on. Hence, the structure and progress varied in each course. The students in the course are evaluated by each teacher with the final grades A, B, C, D, F, based on final exam and several learning activities including quiz and reports. The number of students who obtained each final grade is shown in Table 3.

<table>
<thead>
<tr>
<th>Course No.</th>
<th>Year</th>
<th>Term</th>
<th>Teacher</th>
<th>The number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2015</td>
<td>Spring</td>
<td>A</td>
<td>130</td>
</tr>
<tr>
<td>2</td>
<td>2015</td>
<td>Spring</td>
<td>A</td>
<td>209</td>
</tr>
<tr>
<td>3</td>
<td>2015</td>
<td>Autumn</td>
<td>B</td>
<td>81</td>
</tr>
<tr>
<td>4</td>
<td>2015</td>
<td>Autumn</td>
<td>B</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>2016</td>
<td>Spring</td>
<td>A</td>
<td>161</td>
</tr>
<tr>
<td>6</td>
<td>2016</td>
<td>Spring</td>
<td>A</td>
<td>236</td>
</tr>
</tbody>
</table>
Table 3: Frequency of final grades obtained in the six courses.

<table>
<thead>
<tr>
<th>Grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>673</td>
<td>157</td>
<td>69</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

3. Method

3.1 Recurrent Neural Network

A recurrent neural network (RNN) is a variant of neural networks that handles time series data. An RNN has been applied in the various fields, such as speech recognition and machine translation. In Figure 1 (a) shows a graphical illustration of a structure of an RNN. By inputting data to an RNN, an output value corresponding to the input value is obtained through a hidden layer. At this time, the internal information of the hidden layer based on the past data is input into an RNN, together with the information of input of the present time. Thus, it is possible to output in consideration of the past state.

Figure 1 (b) shows the unfolding in the time of the computation of an RNN. Since the information of the hidden layer at time t-1 is propagated to the same network at time t, an RNN theoretically can output with consideration of all the past information. We can select a method to construct hidden layers, such as Long Short Term Memory and Gated Recurrent Unit (GRU), depending on the way of consideration of the past information. In this paper, we deploy GRU. For the details of RNN, see Bodén, M. (2002).

3.2 Prediction of Students’ Final Grades

A vector of nine kinds of points for each week, that is, an ALP (introduced in Section 2.1) of a student is input into the RNN for each time. The final grade A, B, C, D, or F of the student is considered as the output. Let a number of GRUs included in a hidden layer be 32. The time series data of vectors of nine kinds of points is fed into the RNN, and in each time, the final grade is predicted by the trained RNN. For the training of the RNN, we apply the Back Propagation Through Time (BPTT) to repeatedly update parameters of network and learn optimal parameters.

We also examined the prediction of final grades using multiple regression analysis. For this aim, the final grades A, B, C, D, and F are replaced with 95, 85, 75, 65, and 55, respectively. For each week, the multiple regression analysis using the training data is performed. Then, the final grade is predicted as the final grade whose corresponding value is the closest to the value obtained by applying the multiple regression equation to the test data.

4. Experimental Results

4.1 Prediction by Recurrent Neural Network
We evaluate the prediction performance of the proposed method where, the data on courses conducted in 2015 (Course 1, 2, 3, and 4) is used as the training data, while the data on courses in 2016 (Course 5 and 6) is used as the test data. The numbers of the training data and test data are 540 and 397, respectively. Table 4 shows the accuracy of RNN predictions for each week of the experiment. In the eighth week, the prediction has an accuracy of 84.6%. We also show the accuracy of prediction by the RNN for each final grade in Table 4. We can see that the accuracy of prediction for the final grade A is very high, e.g., the 91.9% in the eighth week. On the other hand, the accuracy of prediction for other final grades is approximately 50%. This may be attributed to a bias in final grades because many students obtained’s final grade of A, while the other final grades have few samples. Among them, the prediction for the final grade F is over 50% in the eighth week. As the number of weeks increases, the accuracy of predictions tends to increase.

<table>
<thead>
<tr>
<th>Week</th>
<th>All students</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71.3%</td>
<td>84.0%</td>
<td>12.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>79.3%</td>
<td>93.4%</td>
<td>6.1%</td>
<td>14.3%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>79.8%</td>
<td>90.4%</td>
<td>27.3%</td>
<td>14.3%</td>
<td>12.5%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>84.6%</td>
<td>95.5%</td>
<td>30.3%</td>
<td>28.6%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>84.1%</td>
<td>94.9%</td>
<td>33.3%</td>
<td>28.6%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>84.9%</td>
<td>94.3%</td>
<td>42.4%</td>
<td>42.9%</td>
<td>25.0%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>85.1%</td>
<td>93.4%</td>
<td>42.4%</td>
<td>42.6%</td>
<td>25.0%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>84.6%</td>
<td>91.9%</td>
<td>54.5%</td>
<td>41.2%</td>
<td>50.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: The accuracy of prediction for each final grade by RNN.

### 4.2 Prediction by Multiple Regression Analysis

We also conduct an experiment using multiple regression analysis. The data on courses held in 2015 and 2016 are used as the training data and test data, respectively.

Figure 2 shows the accuracy of prediction when applying the method using multiple regression analysis to the training data and test data for each week, along with the results by RNN. Here, we also show the accuracy of predictions for each final grade A, B, C, D, and F. In the fourth week, the accuracy for all grades is 64.7%. From the fifth week, the accuracy exceeds 80% and very close to the accuracy of predictions by RNN. As the case of RNN, the accuracy of prediction for the final grade A is high, but the accuracy for other grades is low. This tendency seems to be stronger in multiple regression analysis than with the RNN. In fact, the accuracy of predictions for grades B, C, D, and F is less than 33.3% in the eighth week. Especially, the accuracy of prediction for a grade F is 50% in the case of RNN in the eighth week, whereas it is 0% for each week by the multiple regression analysis.

![Figure 2. The accuracy of prediction by multiple regression analysis and by RNN.](image-url)
5. Discussion

5.1 Consideration of the Results

From the experimental results, we can say that the accuracy of the prediction of final grades by RNN improves as the courses progresses and the data for training and predicting increases. Comparing the results by the RNN and the case of multiple regression analysis shows that the accuracy of predictions after the fifth week is almost same. However, the accuracy by the RNN before the fourth week by RNN is higher than that of multiple regression analysis. Hence, the RNN method is effective in the early prediction of final grades, which is important for both students and teachers. Moreover, the accuracy of prediction for other grades than the final grade A by RNN is generally higher. Table 3 implies that the final grade with a relatively small amount of data can be predicted by RNN. Regarding the final grade F, in particular, the accuracy of prediction is always 0% in regression analysis, whereas RNN achieves 50% accuracy in the seventh and eighth weeks.

In the experiment of this paper, we used the data on the four courses as the training data, where the two courses were taught by Teacher A, and the other two courses by Teacher B. On the other hand, the two courses treated as the test data are conducted by only Teacher A. The difference between the sets of teachers in the training data and test data can be the reason for the decrease in accuracy of predictions.

5.2 Finding the Important Learning Activities which Distinguish the Final Grades

The learning activities including in Active Learner Point that contribute to obtaining a certain final grade can be inferred from the weight of the obtained RNN. First, by observing the values of weight between each input and each GRU of a hidden layer, we can determine which learning activities are important for a unit due to firing. Similarly, by observing the values of weight between each GRU of a hidden layer and each output, we can also determine which units’ firing is important for predicting the final grades of students. By combining these two observations, the items of input that are important to obtain a certain final grade can be inferred.

In our experiment, the following are implied from the obtained RNN:

- To obtain the grade A, a submission of report is important.
- To obtain the grade B, the number of actions in BookLooper and a quiz score are important.
- To obtain the grade C, the logs of Moodle and BookLooper are emphasized evenly.
- To obtain the grade D, the number of markers in BookLooper is important.
- To obtain the grade F, It is important that the number of memos in BookLooper is small.

6. Conclusion

In this paper, we presented a method to predict students’ final grades using a recurrent neural network (RNN). For this purpose, we collected the learning logs from 937 students who attended one of the six courses shown in Section 2.2 for eight weeks. Two different teachers conducted these courses basically following the same syllabus. The nine selected learning logs stored in M2B system are evaluated from 0 to 5 points for each student in each week of the course, and the obtained vector of these nine evaluations is called Active Learner Point (ALP). The ALPs of students and the final grades are treated as the input and output, respectively, of the RNN.

There remain many issues to be investigated along the research direction presented in this paper. Points of particular importance include the following:

- Confirming that the prediction method shown in this paper is effective even when the set of teachers for the training data is completely different from that for test data. This enables us to help teachers who have not yet conducted a course.
It may be valuable to determine whether giving students feedback about the results of prediction, i.e., their potential final grades, in the early stage of a course, would influence their learning activities.

Acknowledgements

The research results have been achieved by “Research and Development on Fundamental and Utilization Technologies for Social Big Data” (178A03), the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan.

References


A Case Study of Interactive Learning Environment for Building Structure of Arithmetic Word Problem in Language Delay

Sho YAMAMOTO\textsuperscript{a} & Tsukasa HIRASHIMA\textsuperscript{b}
\textsuperscript{a}Faculty of Engineering, Kindai University, Japan
\textsuperscript{b}Graduate school of Engineering, Hiroshima University, Japan
*yamamoto@hiro.kindai.ac.jp

Abstract: In special classroom, students are taught solution methods of arithmetic word problems carefully, and the teacher doesn't teach highly effective learning like problem-posing because it is said that its learning load is high for students in special classroom. On the other hands, we tried to decline the cognitive load of problem-posing by analyzing the cognitive load of problem-posing and student’s disability. And then, we developed learning environment based on this analysis. As a result, students with language delay can learn by problem-posing. Therefore, we assumed that it could realize a learning by building structure which is considered to be more advanced and difficult. In this paper, developed learning environment that can build the structure of arithmetic word problem and its experimental use are reported. As a result, students can exercise by building structure and improve their problem-posing performances.

Keywords: language delay, problem-posing, arithmetic word problem, information structure

1. Introduction

Students belonging to special classroom have some disabilities. For examples, disabilities are learning disabilities (e.g. dyslexia), communication disorders, emotional and behavioral disorders (e.g. ADHD) and so on. Therefore, students of special classroom learn an arithmetic word problem by exercises conducted in general classes such as problem solving through their disabilities (William, 2007). Moreover, generally, a student in special classroom is supported to spend their everyday life without problem (Fernández-López et al., 2013). So, their learning progress is slow.

Of course, the teacher would like to do the same teaching as the general classroom if they can do, and they would like to do if they can provide high quality learning. However, because higher loads are generated for more advanced learning (e.g. problem-posing), the teacher judges that it is difficult for students in special classroom who are difficult to learn by problem solving.

We model an arithmetic word problems and develop a learning environment by constructing arithmetic word problems (Hirashima et al., 2007; Yamamoto et al., 2012). For example, it is a learning environment for problem-posing and building structure learning. These are highly difficult learnings that special classroom’s students cannot learn. However, we analyzed various cognitive loads of learning and disabilities of learners in special classroom. And by declining the load related with the disabilities by a system, they were able to learn by problem-posing using our system (Yamamoto and Hirashima, 2016; Yamamoto et al., 2016).

In this study, we analyzed the various cognitive loads and targeted disability, and we assumed that it could realize more advanced building structure learning than the problem-posing, so we carried out experimental use in the special classroom and analysis of the result.

2. Targeted Language Delay and Our Method

The object of this research is a student who is difficult to understand language (e.g. Language Delay). Generally, when understanding sentences, we first divide the target sentences into meaningful chunks.
Then, after dividing the sentences into minimum units such as words, compare it with the mental dictionary, we understand the meaning of the whole sentence (Coltheart et al., 2001; Perry et al., 2007). Therefore, it becomes difficult to divide sentence into meaningful chunks as long sentences, making it difficult to understand sentences.

This difficulty is remarkable for students who make difficulty understanding sentences. For example, several students cannot understand if two or more sentences such as "there are three apples" are unified, and some students cannot understand this sentence as mentioned above. Because they are difficult to understand such sentences, they have to learn an arithmetic word problems by solving problem by shortening sentences or giving pictures by teachers. Other researcher suggested that the learning environment to help these learner by using dynamic visual tool when solving mathematics problems (Peltenburg et al., 2009). And such language delay also interferes with everyday life, so it was necessary to learn about daily life as well. Therefore, their learning progress is very slow.

Cognitive load theory which clarifies the load at the time of learning are proposed (Sweller et al., 1998). In cognitive load theory, the three cognitive loads are defined: extraneous load, intrinsic load and germane load. Here, we consider only extraneous load. Extraneous load is not concerned by learning and decline the amount of resources available to process the intrinsic and germane load. Therefore, it is said that it is a load to be declined. Problem solving and problem-posing in arithmetic word problem are “word problems”, so learners must read or write sentences by themselves. However, reading comprehension and sentence description are loads unrelated to learning an arithmetic word problems. In particular, the description of sentences is far more difficult than reading comprehension. Therefore, due to the load not related to learning, the learner cannot learn by problem-posing.

Our problem-posing are required learner to pose problem by selecting and arranging given simple sentence cards (Yamamoto et al., 2012). Our target arithmetic word problem can be solved with one-step addition or subtraction. This is going to be described in the next chapter. By using this type of problem-posing, we assumed that learners can learn arithmetic word problem through problem-posing by selecting and arranging simple sentences if learners can read even simple sentences given by our learning environment. And previous research showed that this learning environment realized learning by problem-posing and learning effect (Yamamoto and Hirashima, 2016; Yamamoto et al., 2016).

3. Learning by Building Structure and Reading Disability

3.1 Structure of Arithmetic Word Problem

In this research, we targeted an arithmetic word problem which can be solved by one-step addition or subtraction. This structure is explained with reference to Figure 1. This arithmetic word problem consists of three simple sentence expresses quantitative concept. These sentence cards contain quantity, object and attribute. For example, in first sentence, the quantity is five, the object is apple, attribute is “there are”. Attribute shows the kinds of quantity that are independent quantity expresses existence of quantity, and relative quantity expresses relation between other existence quantities. For example, the third sentence contains the attribute that is “altogether”. This attribute expresses the relation between apple and orange. The story of arithmetic word problem is decided by relative quantity sentence, where are combine, change-increase, change-decrease and compare. We call this model as triplet structure model (Hirashima et al., 2014). Also, the difference between story and problem is the given three simple sentence cards include required value or not. In our problem-posing, learner gives a calculation and the story and then he/she is required to pose the problem by selecting and arranging a given sentence cards.

The relation of these quantities is shown in Figure 2. We call this expression part-whole relation and the block sets the simple sentence cards called Tape-Block. Above part of Tape-Block expresses the whole quantity, for example, the sentence of apple and orange. Below parts of Tape-Block expresses the part quantity, for examples, the sentence of apple and the sentence of orange. The relation between three quantities in arithmetic word problem are visualized by this model in each kind of story. Therefore, this kind of arithmetic word problem include three numerical relation, where one addition and two subtraction. In Figure 2, there are three numerical relation that are “8-5=?”, “8-?=5” and “5+?=8”. We called this relation the one addition and two subtraction relation. This relation is different in each story.
There are two kinds of numerical relations in the arithmetic word problem that can be solved by one-step addition or subtraction. The story of this arithmetic word problem is divided into addition story and subtraction story. Addition story is usually expressed by combine story or change-increase story. Subtraction story is usually expressed by change-decrease story or comparison story. So, the story of Figure 1 is addition story. Therefore, the numerical relation of this problem expresses as “$5+?=8$” because the story of Figure 1 is addition story. We call this numerical relation the story numerical relation. On the other hand, we are able to solve this problem by “8-5”. We call this numerical relation the calculation numerical relation. In this problem, story numerical relation and calculation numerical relation are different. We call this kind of problem "reverse thinking problem". Reverse thinking problem is much harder than "forward thinking problem" where story numerical relation and calculation numerical relation are the same ones.

Figure 1. Example of Problem by Triplet Structure Model.

![Figure 1](image)

Figure 2. Part-Whole Relation on Tape-Block and Three Numerical Relation of Figure 1’s Problem.

3.2 Relation Between Structure of Arithmetic Word Problem and Reading Delay

In previous research, it was found that students who have difficulty understanding language can achieve learning with a high-level difficulty by declining the load that related to language understanding and included in extraneous load. In other words, if their disability is the same as the load included in extraneous load, advanced learning can be realized. Therefore, in this research, we assumed that we could develop a learning environment that realizes the exercise of building above structure more difficult and effective than problem-posing.

If learners would like to think about problem structure, the visualization and building structure are effective (Hirashima and Hayashi, 2016). Therefore, it is the purpose of building structure learning to operate and understand the triplet structure model and the one addition and two subtraction relations of Figure 2 described in the previous section. The activity is done by the learner is mainly the problem-posing as sentence integration and the simple sentence setting to Tape-Block. By analyzing the cognitive load of this exercise, the cognitive load of problem-posing is the same as that of the previous research. It is necessary for the cognitive load of building structure to read and understand the simple sentence cards, and understand Tape-Block expression. Reading and understanding of the simple sentence card are a load that is required even in problem-posing as sentence integration, so a target learner can read a simple sentence. In addition, since Tape-Block is a graphical expression and does not involve sentence comprehension, we thought that the target learner can understand this expression. So, we performed experimental use of its learning environment that realizes the building structure exercise.

4. Interactive Learning Environment for Building Structure: MONSAKUN Tape-Block

Figure 3 is interface of learning environment for building the arithmetic word problem structure called MONSAKUN Tape-Block. First, learners log in this system by selecting their class and grade. Next, the learning environment displays the interface of level selection and then, learners selects one of levels one...
to ten in this interface. The assignments of this learning environment are omitted for the page limited but it is designed to gradually learn the structure. When learner selects any level, our learning environment shows the interface shown in Figure 3.

Interface for problem-posing shown in Figure 3 (a) gives learner the assignment and several sentence cards. Learner can pose the problem by selecting three sentence cards from given ones and arranging them in proper order. If the posed problem is satisfied given calculation and story, MONSAKUN Tape-Block feeds back the correct and shows next assignment. Through this exercise, learner can learn the triplet structure model. On the other hands, interface for building part-whole relation is shown in Figure 3 (b). This interface is required learner to build a part-whole relation by setting given sentence cards to Tape-Block to satisfy given calculation and story. By this learning, learner understands about the part-whole relation. These interface feedbacks two feedbacks. One feedback is flag feedback that replies only correct or not. Other one is pointing hint that points learner’s attention to errors.

(a) Problem-posing (b) Building Part-Whole Relation

Figure 3. Interface of MONSAKUN Tape-Block for problem-posing.

5. Experimental Use

5.1 Subjects and Procedure

The subjects were thirteen students at special classroom in junior high school. They have already finished to learn the arithmetic word problems that can be solved by one-addition or subtraction. There are a few subjects because students at special classroom in Japan are small. We divided them into following three groups. Four subjects don’t understand the simple sentence but they can read simple sentence (group A). Four subjects understand and read the simple sentence but they cannot read long sentence that is combined more than two simple sentence (group B). Five subjects understand long sentence (group C).

We use two our learning environments that are learn by problem-posing and building problem structure. There are called MONSAKUN Touch 1 (MT1) and MONSAKUN Tape-Block (MTB). If learner not enough to understand the problem structure, he/she cannot pose the problem by MT1. So, we use the MT1 for verifying the understanding of problem structure in each subject. In this experimental use, first, a subject exercise by MT1 as pre-test in one lessons. One lesson is forty-five minutes. Second, a subject learn s by MTB in three lessons. Subjects are taught the method of each exercise at twenty minutes in first lesson. They exercise by MTB at remaining twenty-five minutes. Finally, they exercise the problem-posing by MT1 in one lesson.

Four teacher in special classroom and one teacher teach mathematics are join in experimental use. They evaluated the subjects are not able to exercise and learn by MTB before we perform experimental use because learning problem structure is very high-level learning for students in special classroom. However, teachers would like to realize the learning problem structure of the student can. So, we suggested the learning method by visualizing and building the structure of arithmetic word problem. We also assumed that (a) Subjects who can understand the simple sentence are able to exercise the learning by problem structure, (b) Subjects who can understand the simple sentence are able to improve their problem-posing performance.
5.2 Results

We report the results of experimental use that are a result of MTB, MT1. We couldn’t perform the statistical analysis because the number of subjects are small. First, we describe about the result of MTB by Table 1. All subjects concentrated to exercise during each lesson. Group A achieved level 7, Group B and C achieved level 10. Average of accuracy rate is sixteen percent in group A. Average of accuracy rate of group B is forty-five percent, group C is seventy-one percent. Therefore, all subjects can exercise the learning of problem structure by MTB but group A is difficult to learn.

The result of MT1 is shown in Table 2. This table shows the average correct number of problem-posing in MT1 in each kind of problem. The assignment of reverse thinking problem-posing is most difficult in problem-posing exercise and the forward thinking (forward calculation) is most easy. All subjects can pose the problem by MT1 and they improve their performance of problem-posing in reverse thinking problem and total. There is an approached significant between the average number of correct problem in pretest and posttest (Paired t-test, \( p=.07<.1 \)). There is a significant difference between the average number of correct reverse thinking problem in pretest and posttest (Paired t-test, \( p=.02<.05 \)). Next, we analyzed the data of reverse thinking problem in each group. The correct number of group A didn’t increase because they didn’t learn by MTB enough. The correct number of group B increased a lot. This result suggested that the learner of group B has be able to exercise and learn the problem structure by using MTB. In group C, the correct number increased so this group also has been able to exercise and learn the problem structure by MTB.

### Table 1: Average Accuracy Rate of MONSAKUN Tape-Block in Each Level (MAX: 1).

<table>
<thead>
<tr>
<th>Group</th>
<th>Lv1</th>
<th>Lv2</th>
<th>Lv3</th>
<th>Lv4</th>
<th>Lv5</th>
<th>Lv6</th>
<th>Lv7</th>
<th>Lv8</th>
<th>Lv9</th>
<th>Lv10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.42</td>
<td>0.28</td>
<td>0.23</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0.54</td>
<td>0.40</td>
<td>0.48</td>
<td>0.36</td>
<td>0.52</td>
<td>0.67</td>
<td>0.71</td>
<td>0.36</td>
<td>0.3</td>
<td>0.14</td>
</tr>
<tr>
<td>C</td>
<td>0.84</td>
<td>0.78</td>
<td>0.86</td>
<td>0.31</td>
<td>0.78</td>
<td>0.74</td>
<td>0.53</td>
<td>0.81</td>
<td>0.74</td>
<td>0.67</td>
</tr>
</tbody>
</table>

### Table 2: Average Correct Number of Problem-posing in MONSAKUN Touch 1.

<table>
<thead>
<tr>
<th>MAX</th>
<th>Forward thinking (Forward calculation)</th>
<th>Forward thinking (Reverse calculation)</th>
<th>Reverse thinking</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>group</td>
<td>pre</td>
<td>post</td>
<td>pre</td>
</tr>
<tr>
<td>A</td>
<td>11.2</td>
<td>9.8</td>
<td>8</td>
<td>9.8</td>
</tr>
<tr>
<td>B</td>
<td>12</td>
<td>12</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>11.75</td>
<td>12</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>ALL</td>
<td>11.62</td>
<td>11.15</td>
<td>15.08</td>
<td>16.08</td>
</tr>
</tbody>
</table>

5.3 Discussion

We reported the results of experimental use in previous section. The result of exercise by MONSAKUN Tape-Block suggested that the learner with reading disability has be able to exercise by building the problem by MONSAKUN Tape-Block. Moreover, the result of pretest and posttest shows the improvement of problem-posing performance in group B and C. So, reading disability can learn the problem structure if they can understand the simple sentence. Learning problem structure is impossible learning in reading disability so this result suggested the realization of more advanced learning.

The subject of group A can exercise the MONSAKUN Tape-Block and MONSAKUN Touch 1 but they couldn’t learn the problem structure. This means that it is effective for them to exercise by visualizing and building the problem structure but kit is not proper. We have to realize the learning environment for building the simple sentence. By this experimental use, we have verified the stage of reading disability. If the learner understands the arithmetic word problem that can be solved by one-step addition or subtraction, they have to understand simple sentence. So, if the learner understands the simple sentence, they can learn the structure of arithmetic word problem even though they have language disability.
In conclusion, if the learner with reading disability is given the proper kit like simple sentence for learning, they can exercise and learn more effective and difficult learning like the structure building. Of course, learning environment is needs to be interactive.

6. Conclusions

In the special classroom, careful learning is carried out due to the student’s disabilities. By considering the cognitive load of learning, we are realizing more advanced learning in the special classroom by supporting only disabled activities in the system. In this research, we tried to realize the building structure learning of arithmetic word problems for students who are difficult to understand sentences. As a result of experimental use, it was found that building structure exercises by our learning environment can be realized for the target learner.

In future works, we improve the interface for reading disability and verified its effect. Also, the confirmation of our assumption in other domain like an arithmetic word problem that can be solved by one-multiplication and division is important.

Acknowledgements

I am deeply grateful to the teacher and the students in the elementary school for performing the experimental use and giving insightful comments. This work was supported by JSPS KAKENHI Grant Number 17K12954.

References


A Learning Support System for Integrated Motor Skill by Organized Training Stages

Kohta SUGAWARAa, Takehiko YOSHIKAWAa, Kenji MATSUURAa*, Karungaru STEPHENa & Naka GOTODAb

aGraduate School of Advanced Technology and Science, Tokushima University, Japan
bInformation Technology Center, Kagawa University, Japan
*ma2@tokushima-u.ac.jp

Abstract: It is effective to use a learning framework when learners acquire a new skill. Jump rope is one of a slightly complex exercise which requires the coordination of each primitive movement and to control a rope. Therefore, this study designs a learning support system using learning stages which shift from primitive training to cooperative training according to learner’s performance. The prototype is equipped with 3 functions which detect the feature points, compare the learner’s data with an expert’s and provide the learner with the visualized information using image processing technique. This paper describes a learning process, methodology and issues encountered.

Keywords: Motor skill, learning support, feedback, image processing

1. Introduction

1.1 Background

Skill learners often need to acquire some techniques using trial and error to execute an unlearned performance. Furukawa (2009) states that it generally is difficult for novices to understand and acquire motor skills. One of the reasons is that some novices make mistakes while executing and controlling their bodies by themselves due to the mental model error. This is one of the problem which makes it difficult for some novices to acquire motor skills. Therefore, this study develops a learning system to support them acquire a skill.

Guthrie (1952) defines a skill as an ability that is learned to reach a decided achievement in advance. It is often accompanied by maximum accuracy while consuming minimum time/energy or both minimum time and energy. Motor skill is a skill in which people execute performance moving their body. For example, in swimming, it is required to move legs, arms, and breathe simultaneously. To acquire the skill and swim smoothly, one needs to learn and train these skills, at first, independently, after then, cooperatively. This study suggests learning stages when novices acquire such skill.

1.2 Jump Rope

Jump rope is one of the well-known exercise all over the world. The characteristic is that anyone can do it alone and it doesn’t need a wide space. The notable feature is that combining several motor skills and changing the activity degree can produce many tricks. Gotoda et al. (2010) classified these tricks into two groups; rotate the wrist once per jump (single bounce) and rotate the wrist twice per jump (double bounce). The difference between two groups is the degree of difficulty. Some people who can do the single bounce cannot do the tricks of double bounce. Hence, our supporting targets are those who cannot execute double-under, the most fundamental trick in double bounce.

Fitts et al. (1967) claim that it is important to consider three learning stages; verbal-cognitive, motor and autonomous stages. At first, in verbal-cognitive stage, the problem for a learner is the language and cognition. In case of jump rope, the important things are how to decide the goal, evaluate the performance, control a rope, and pay attention. Secondly, in motor stage, since learners have already solved most of cognitive problems, they focus on more effective organization of motion pattern. In case
of hand motion in double-under, learners begin to make an exercise program for rotating the wrist more rapidly than single-under (Yoshioka et al., 2015). By this trial and error, they become able to detect their own error and get self-feedback, hence, their performance is improved rapidly. Finally, autonomous stage means the development of autonomous motion which does not need paying attention to. In case of double-under, one can continue to repeat a set of motion units as long as one has the physical strength. Therefore, this study focuses on novices who belong to the verbal-cognitive or motor stage.

2. Strategy for Jumping Rope

2.1 Primitives and its Relations

Jump rope exercise mainly consists of hand and jump motions. ARSA, which stands for Australian Rope Skipping Association, suggests how to execute the tricks in jump rope for a learner. In double-under instruction by ARSA, there are three primitives of hand motion and three primitives of jump motion. Primitive movements of hand are “rotating the wrist”, “quick circular” and “beat of 1, 2”. In addition, primitive movements for jump are “jumping high”, “not bending at the waist” and “straight knees”. Based on these primitive movements, this study examines and considers learners actions.

2.2 Training Stages for Acquisition

This study also interprets from ARSA’s suggestion that, there are 4 training stages to acquire double-under. Stage 1 and 2 are basic stages where learners acquire primitive movements independently without a rope. Stage 3 is an advanced stage where they learn to coordinate primitive movements by combining hand motion with jump motion. In this stage, they practice without a rope. Here, considering time course in jump rope exercise, it is an exercise which repeats a set of motion-unit. The goal of the repetitive exercise is to continuously perform. However, if novices practice with a rope at first, it arises a problem which they cannot continue because they have not acquired motor skills for the trick yet. This can lead to failure when they begin to execute. Therefore, in these stages, one of the purpose is that they acquire a form to execute in double-under since they can repeat to execute without a rope. Stage 4 is also an advanced stage where learners try to coordinate primitive movements with a rope.

Meanwhile, this study points at a problem between stage 3 and 4. The difference between them is whether a rope is used or not. However, there is a case that the form of some novices become worse when they have passed stage 3 and try to practice stage 4 at first. One of the reason is that they cannot control a rope. In stage 3, the goal is to acquire the form in double-under without a rope, not to make a model to control with a rope. This study inserts a new stage 3’ before proceeding to stage 4. It is also an advanced stage where they learn to coordinate primitive movements with a disticketed rope. The idea is that they execute double-under with a rope which is rotated on only one side of their body. In other words, a rope doesn’t need to pass under their foot. The important thing is that they need to achieve the trajectory of a rope’s top observing from their side as drawn like a circular orbit by rotating the wrist. Hence, this study construct 5 training stages by inserting stage 3’.

2.3 Transferring between Stages

It is required to learn how to transfer between stages. This study defines the stage flow based on ARSA. Figure 1 indicates a relationship between training stages. In each stage, there is a function’s cycle which transfers from observation to data analysis, stage’s decision and suggestion in a system. At first, a learner tries to execute double-under with a rope and without it as a pre-trial, then the system observes, judges and determines the stage. Here, there are some rules to transfer to the other stage. A learner, at first, tries to practice on the stage once at the first stage or the transferred stage, then the system judges the performance and determines the next stage to practice.
Next, this study proposes how to judge each stage. In stage 1, 2 and 3, the system judges by binary method which determines if the learner’s data is good or bad per a primitive movement based on reliable ranges calculated by expert’s data statistically. In stage 3’, the system again uses the binary method to decide whether the rope passes under one’s foot while jumping or not. In stage 4, it is difficult to observe the performance because it is not clear if the learner executes double-under or not.

In addition, this study proposes how to transfer to the other stages. Basically, the system transfers from stage 1 to stage 4 in order. However, if a learner cannot master the skill in the stage, it is possible to train in the same stage again or transfer to lower stage. In case of stage 3, Sugawara et al. (2016) indicate that the coordination of primitive movements is required. That is to say, a learner who has been trained in stage 1 and 2 acquires the coordination which consists of primitives in stage 3 although some primitive movements become bad. In addition, if the judgement of the coordination is correct, the system transfers to stage 3’. If not, the system adopts in stage 3 again. In stage 3’, if the judgement is that a rope passes under one’s foot on all jumps while a learner execution is correct, the system transfers to stage 4. If not and the judgement of the coordination is correct, the system adopts in stage 3’ again. Otherwise, the system stays in stage 3. In stage 4, if a learner can execute double-under at least once within 2 trials, the end of this learning support is reached. If not, the system transfers to stage 3’ and a learner trains at it again.

3. Methodology

3.1 System Work

In learning process, it needs to judge whether a learner has acquired the skill or not per stage. This study mainly tries to solve it by a parameterized performance for the judgement in jump rope exercise. Hence, this study suggests the way to detect feature points using techniques in image processing.

3.1.1 Viewpoint

Learners need to grasp the way of using their bodies by themselves objectively. Jump rope exercise is an exercise which consists of hand and jump motions. In double-under, by observing from the side of a learner, it regards a motion of rotating the wrist as circular motion. In addition, it regards a motion of jump as vertical motion. Therefore, this study uses the way of observation in two-dimensions using a video camera.

3.1.2 Data Analysis

For detecting body information of a learner, this study uses color recognition. Herewith, it gets several points; head, hand, the top of grip and toe. Using these data, this study calculates reliable ranges that it is possible to execute double-under from the data of experts statistically. This study defines experts as those people who can execute double-under over than 20 times, then gathers the data from 10 such experts. Figure 2 shows a reliable range of “rotating the wrist” as an example. We classified 3 groups;
experts could execute it “under 10 times”, “over 10 and under 20 times” or “over 20 times” until failure. Every group consists of over 8 trials data. The results of Figure 2 suggest that those who can execute it more times tend to execute in a smaller radius. However, the target in this study is novices. Therefore, this study adopts the reliable range which experts execute “under 10 times” until failure to judge the novices and similarly adopts the same way for other primitive movements.

![Figure 2. Reliable range of rotating the wrist](image)

### 3.2 Learning Process

To acquire skill, we need to consider learning flow. Figure 3 is a learning process in this study based on Soga et al. (2005). This flow consists of 2 parts; a learner’s part and a system’s part. In learner’s part, there are 3 parts in a cycle; execution which a learner tries to do, recognition which figures out one’s error arising from own movement by feedback and improvement in which one takes a strategy for success and acquiring the skill based on recognition. In the system part, there are 3 functions in a cycle; observation of learner’s execution, data analysis and indication including the result of analysis as a learning support framework. “n” stands for the number of stages.

![Figure 3. Learning process](image)

### 3.3 Timing of Support

For learning support, we need to consider timing when learners get some information. Kosaka et al (2011) suggested the timing of support. In “synchronized support”, learners get support while executing appropriately. The advantage is that learner can fix and improve the motion immediately. The weakness is that a learner cannot get an evaluation of the overall performance. Furthermore, it is not suitable for ballistic exercise. On the other hand, in asynchronous support, the advantage is that a learner can get information of both motion and overall performance. The weakness is that it is possible to lose consciousness in each motion since a learner cannot get some information simultaneously while executing. Therefore, this study adopts “asynchronous support” as the learning support. The reason is that the goal is to acquire double-under skill which consists of some primitive movements and their coordination. Therefore, we need to provide a learner with the information and whole body movement.

### 4. Designing Issue
4.1 Observation

This study uses OpenCV as techniques of image processing for detecting the feature points. However, it is difficult to track the top of a rope by color recognition because of the speed and the environment. Karungaru et al. (2016) propose a method using particle filter for tracking the feature. Therefore, this study applies color recognition for detecting the feature points, in addition, background subtraction and particle filter for tracking the features.

4.2 Learning Support

It is effective for learning support to consider feedbacks. According to Schmidt (1991), feedback often means information difference between the target value and performance. There are two feedbacks in detail; intrinsic and augmented. The former is the information in which learners recognize their performance by themselves while executing, in case of jump rope, they often get it by visual and auditory sense. The latter is the information which is based on artificial methods, for example coach’s voice, observed data and movie. Some novices make miss-recognition in the intrinsic one. In contrast, augmented feedback is adjusted accurately and objectively. This is information that coaches don’t or cannot control. Hence, this study designs augmented feedbacks for a learning support system. Therefore, this system is equipped with visualized feedback. According to Yoshioka et al. (2015), a learning system supplies a learner’s current movie attached with target form about hand motion. The purpose is that learners become able to imagine the target model compared with learner’s internal model by themselves. The system contains a video of actual performance attached the target points and a model video made by experts. Figure 4 is one of the sample as a feedback video in stage 3’. Stage3’ is for those who cannot rotate the wrist twice while jumping with a rope in spite of acquiring ideal form without a rope for double under in stage 3. The video indicates a trajectory (two red circles) and the speed of rotation (a light blue dot moving on the trajectory) to achieve an ideal rotation with a rope while jumping. The target movements are overlapped on learner's result and drawn by image processing. Learners watch it and are given any target values, after that, they make a strategy with moving their bodies by themselves for the next execution.

Moreover, in stage 3 and 3’, the coordination between primitive movements is required. The system needs to consider the correlation between them in feedback. Therefore, this study calculates the correlation between them from the result of 10 experts. In Table 1, (h1) and (h2) mean primitives of hand motion (this study treats the parameter of “beat of 1, 2” as binary, then excludes it from this calculation), and (j1), (j2) and (j3) mean primitives of jump motion. In addition, the correlations are combined hand motion with jump motion.

<table>
<thead>
<tr>
<th>Serial number</th>
<th>combination</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(h2) – (j1)</td>
<td>0.881210473</td>
</tr>
<tr>
<td>2</td>
<td>(h1) – (j3)</td>
<td>-0.525911049</td>
</tr>
<tr>
<td>3</td>
<td>(h1) – (j1)</td>
<td>0.35485142</td>
</tr>
<tr>
<td>4</td>
<td>(h2) – (j2)</td>
<td>-0.335502835</td>
</tr>
<tr>
<td>5</td>
<td>(h1) – (j2)</td>
<td>0.221594898</td>
</tr>
<tr>
<td>6</td>
<td>(h2) – (j3)</td>
<td>-0.057921484</td>
</tr>
</tbody>
</table>

Figure 4. The feedback video in stage 3’
This study uses these results for a decision of a training way. It expects that a learner in stage 3 or 3’ needs to acquire some skills. Although it is not effective for people to train all skills simultaneously, a learner needs to train some skills simultaneously to acquire the coordination. This system decides a training method which a learner should train at first. For example, if (h2) is judged as bad, this study chooses the combination between (h2) and (j1) from 3 choices because we assume that the correlation is higher. Finally, the system provides a learner with a training way for improving the combination.

5. Conclusion

This study proposes the framework of learning support to acquire motor skill in double-under. To achieve this, the study configures learning stages from fundamental training to complex one based on ARSA. Since it is important to connect between the stages, this study adopts the judgement which is determined statistically by the data of learner’s performance. Accompanied with this, this study develops the system which is equipped with three functions; observation using image processing, data analysis and indication as feedback. In observation, this study tries to use the technique of color recognition and background subtraction for detecting the body information, and particle filter for tracking feature points. In feedback, this study supplies a learner with the visualized information which is a learner’s movie and a model video made by experts for their understanding objectively. In addition, the timing when the feedback is provided to a learner is an asynchronous support. The next issue using this prototype is an evaluation whether the learning framework is effective for novices or not.

Acknowledgements

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP15K01072. The system development was supported by an alumnus Mr. Yoshioka.

References

Enhancing Students’ Critical Reading Fluency, Engagement and Self-Efficacy using Self-Referenced Learning Analytics Dashboard Visualizations

Christin JONATHAN*a, Jennifer Pei-Ling TANa, Elizabeth KOHa, Imelda CALEONa & Siu Hua TAYb

aNational Institute of Education, Nanyang Technological University, Singapore
bEducational Technology Division, Ministry of Education, Singapore
*christin.jonathan@nie.edu.sg

Abstract: Although learning analytics (LA) dashboard visualizations are increasingly being used to provide feedback to students, literature on the effectiveness of LA dashboards has been inconclusive. To address this, a LA student dashboard visualizing students’ latest data against their own data from previous weeks (i.e., self-referenced data) was designed – informed by Fredrickson’s (2004) broaden-and-build theory, as well as studies highlighting personal best goals (Martin & Elliot, 2016) and the negative effects of peer comparisons (Corrin & de Barba, 2014). The self-referenced LA student dashboard was implemented and evaluated in a Singapore Secondary school as part of a larger study, WiREAD. This paper reports on the quantitative impact of the WiREAD self-referenced LA dashboard visualizations on 15-year-old students’ critical reading fluency, cognitive reading engagement, and English language (EL) self-efficacy, as well as students’ qualitative feedback on the usefulness and shortcomings of the LA dashboard.

Keywords: Learning Analytics, student dashboards, English language, critical reading skills

1. Introduction

Learning analytics (LA) dashboards are increasingly being employed to provide timely, dynamic, and visual online formative feedback to support teaching and learning (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013) and make learning visible through the use of visualizations (Lockyer, Heathcote, & Dawson, 2013). Research studies in the field of LA have highlighted the potential of LA student dashboards in enhancing students’ motivation and engagement (Verbert et al., 2013; Wise, Zhao, & Hausknecht, 2014), as well as improving learning behaviours and academic performance (Arnold & Pistilli, 2012). However, some have suggested that LA dashboards have resulted in more frequent but not higher quality feedback (Pardo et al., 2016; Tanes, Arnold, King, & Remnet, 2011) and can even be detrimental for learning (Corrin & de Barba, 2014). These contradictory illustrations of the effects of LA dashboards on learning emphasize the need for the purposeful and empirically-informed design of LA dashboards.

The use of LA student dashboards visualizing peer comparisons such as the class average can dampen the engagement of students above the class average (Corrin & de Barba, 2014) and be perceived as discouraging and stressful (Wise et al., 2014; Tan, Koh, Jonathan, & Yang, 2017). In addition, personal best goals have recently been given prominence as strong predictors of academic motivation and engagement (Martin & Elliot, 2016). According to Fredrickson’s (2004) broaden-and-build theory, positive emotions serve to broaden mindsets and responses that build up lasting personal resources associated with greater student engagement and resilience in school settings (Reschly, Huebner, Appleton, & Antaramian, 2008). Taken together, these findings suggest that LA student dashboards should be designed to be more intrinsically motivating through the use of an individual’s own learning data as the comparison for their performance (i.e., self-referenced data).
To this end, a self-referenced LA student dashboard visualizing students’ learning data from past weeks against their latest data was designed and implemented as part of a larger ongoing study — WiREAD. This paper reports on the effectiveness of the WiREAD self-referenced LA student dashboard in enhancing 15-year-old students’ critical reading fluency, cognitive reading engagement and EL/reading self-efficacy in English language in a Singapore Secondary school.

2. The Current Study

WiREAD is a web-based collaborative critical reading and LA environment designed with the aim of fostering students’ 21C literacies in EL, particularly their critical reading development. Drawing from theorizations of new literacies (Jewitt & Kress, 2003; New London Group, 1996) that view literacy as “increasingly multiple, multimodal and mediated through new technology” (Burnett, Davies, Merchant, & Rowsell, 2014, p. 1), our conceptualization of 21C literacies emphasizes the importance of critical reading as a key component of language and literacy skills required for individuals to thrive in modern knowledge-based societies. Through WiREAD’s collaborative critical reading space, multimodal texts around social, moral, and/or ethical dilemmas were uploaded for students to read and critique with their classmates during and beyond formal class time (see Tan, et al., 2017). Students’ activity on WiREAD was captured and visualized to them through the LA dashboard, along with other relevant learning data. We elaborate on the components of the LA student dashboard in the following section.

2.1 Self-referenced LA Student Dashboard Visualizations

The WiREAD LA student dashboard was designed to improve learning outcomes by enabling students to track their critical reading engagement and progress, and change their learning behaviours and strategies. Students were given access to visual, dynamic, and timely formative feedback around their discourse, dispositional, and social network analytics (Ferguson & Buckingham Shum, 2012) as well as achievement data. Building on a previous version of the LA dashboard (see Tan et al., 2017), the visualizations were modified in this iteration to show students their latest available learning data compared against their past data (i.e., self-referenced data). This self-referenced LA student dashboard is the focus of this paper, and comprises the following 5 components (Figure 1):

- **My Comments and Replies Data**: discourse-related learning data visualising students’ latest and previous week’s frequency and length of comments and replies posted on each text, alongside the number of peer-awarded ‘likes’ and teacher-awarded motivational badges received;
- **My Critical Lens and Comment Types Usage Data**: discourse-related learning data visualising students’ latest and previous week’s frequency of the critical lens and comment types that students had to tag to each of their comments/replies for each text;
- **My WiREAD Social Learning Network Map**: social network maps visualising the position and influence of students in the WiREAD learning network for each text (only the names of people the student has discussed the texts online with are visible);
- **My Learning Attitudes and 21CC Profile**: dispositional learning data visualising students’ responses on self-report questionnaires measuring 21C learning dispositions at the start and end of the trial;
- **My Achievement Data**: achievement data visualising students’ reading scores from school-based EL assessments that occurred before and during the trial.
3. Methods

A mixed methods research design was used to gather quantitative and qualitative data pre- and post-trial. The sample is made up of 101 students from 7 Secondary 3 (Grade 9) classes who had been randomly assigned the self-referenced LA dashboard during a 10-week trial in Singapore.

Critical reading fluency was measured using an objective critical reading test designed and graded by teachers to assess students’ range of critical reading sub-skills demonstrated in their answers (e.g., agreement/validation, disagreement/challenge, justification). A self-reported questionnaire was used to measure (i) cognitive reading engagement using a 4-item scale adapted from Wolters’ (2004) learning strategies questionnaire (see Caleon et al., 2015), and (ii) EL/reading self-efficacy measured by a 10-item EL academic self-concept scale (adapted from Rosenberg, 1989) and a 13-item critical reading ability scale. These scales demonstrated good internal reliability, with Cronbach alpha scores ranging from 0.84 to 0.96. Sample items of each scale are listed below.
(i) Cognitive reading engagement (7-point Likert scale, 1-strongly disagree to 7-strongly agree): In EL classes, I try to relate what I’m learning to what I already know.

(ii) EL self-efficacy:
- EL academic self-concept (7-point Likert scale, 1-strongly disagree to 7-strongly agree): I feel that I am able to do tasks as well as most other classmates.
- Critical reading ability (7-point Likert scale, 1-never true to 7-always true): I am capable of examining the assumptions underlying the EL texts I read.

Qualitative data was also collected through (1) qualitative feedback forms asking students to elaborate on the usefulness and shortcomings of the self-referenced LA dashboard and (2) student focus groups comprising 8-10 students per focus group per participant class.

4. Results

4.1 Quantitative Findings

Analysis of the quantitative data using paired-sample t-tests revealed the effectiveness of the self-referenced LA dashboard as shown by the statistically significant pre- and post-trial improvements in students’ (i) critical reading fluency scores ($t=2.72$, $p<.01$, $d=0.22$), and self-reported (ii) cognitive reading engagement ($t=2.81$, $p<.01$, $d=0.27$), and (iii) EL/reading self-efficacy, comprising the subscales of EL academic self-concept ($t=2.19$, $p<.05$, $d=0.15$) and critical reading ability ($t=3.27$, $p<.01$, $d=0.32$). The descriptive data, t-test results, and effect sizes are reported in Table 1.

Table 1: Descriptive data, paired-samples t-test results and effect sizes.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Pre-test Mean (SD)</th>
<th>Post-test Mean (SD)</th>
<th>t</th>
<th>cohen’s d (Effect size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical reading fluency</td>
<td>10.23 (3.96)</td>
<td>11.18 (4.66)</td>
<td>2.72**</td>
<td>0.22 (s)</td>
</tr>
<tr>
<td>Cognitive reading engagement</td>
<td>4.95 (1.07)</td>
<td>5.22 (0.90)</td>
<td>2.81**</td>
<td>0.27 (m)</td>
</tr>
<tr>
<td>EL academic self-concept</td>
<td>4.39 (1.10)</td>
<td>4.57 (1.17)</td>
<td>2.19*</td>
<td>0.15 (s)</td>
</tr>
<tr>
<td>Critical reading ability</td>
<td>4.26 (1.09)</td>
<td>4.60 (1.01)</td>
<td>3.27**</td>
<td>0.32 (m)</td>
</tr>
</tbody>
</table>

*p<.05,**p<.01. SD = Standard Deviation.

4.2 Qualitative Feedback

Qualitative data from feedback forms and focus group discussions provided a more nuanced understanding of the usefulness of the self-referenced LA dashboard as experienced by the students, as well as the shortcomings associated with the use of such LA affordances.

4.2.1 Perceived Benefits

Students’ qualitative accounts of the usefulness of the self-referenced LA dashboard visualizations for their learning highlighted three benefits: 1) creating greater self-awareness, 2) encouraging deeper learning engagement, and 3) promoting the development of critical reading fluency.

Students frequently described the LA dashboard visualizations as useful in terms of helping them “track [their] improvement” and become more “aware of their strength and weakness”: “It is useful in a way that it provides a clear view of my progress and my current standing in terms of different components such as visual text and narrative text. It highlights where my strengths and weaknesses are and where I should focus and try to improve on” (Student 3r729). Similarly, Student 3r141 stated that “My Critical Lens/Comment Types... allows me to know which critical lenses I have not used. Without it, I am sure that I would be using the same critical lens repeatedly.”

Repeatedly heard among the students were comments about how the LA dashboard visualizations “motivated” and “encouraged” them to deepen their critical reading engagement: “Having more likes and motivational badges also motivate me to put in more effort into WiREAD to make me use more effort into replying and commenting” (Student 3r713). Echoing this view, Student
3r527 stated that “My Critical Lens/Comment Types... encouraged me to spread out and try different critical lenses and comment types for a more comprehensive view of a topic.”

Many of the students recognised that their learning data, particularly their Critical Lens and Comment Types Usage data, could help in “improving [their] quality of responses”: “I found My Critical Lens/Comment Types presented in My Learning Dashboard very useful because I can see the frequency of me using different critical lens and comment types in the texts I have read. Then I will try to use a critical lens or comment types that I did not use very often next time. By trying different critical lens and comment types, I can develop my reading skills” (Student 3r723).

4.2.2 Perceived Shortcomings

Three key shortcomings of the LA dashboard emerged from students’ qualitative feedback: 1) desire for more system/expert-generated quality indicators, 2) preference for ‘live’ data indicators, and 3) an emergent understanding of the influence of dispositional factors and social learning connections on learning.

Teacher Motivational Badges were implemented on WiREAD instead of the formal grading of comments and replies and were used by some students “to check how well written [their] comments are”. However, some students called for more quality-based indicators of critical reading, stressing that “quality is over “quantity”: “I think the My Comments/Replies section does not really help as it only provides information of how many replies and comments I have made and not the quality of my answers” (Student 3r132).

Many expressed a strong preference for ‘live’ indicators, describing ‘My Learning Attitudes and 21CC Profile’ and ‘My Achievement Data’ visualizations as “a one-time check” that “does not reflect the improvements [they] had made”: “the 21CC Profile was created based on a 1 time survey and even if we feel different, we are unable to change it and it just remains permanent. Hence, I feel that we should be able to change it every month to see if there is a difference” (Student 3r714).

Most students displayed an emergent understanding of the significance of dispositional factors and social learning connections as they “do not see why [their] interactions with other users has any impact on [their] learning” and felt that data on learning attitudes and 21CC profiles “does not have any effect on [their] learning”. Views such as “my WiREAD Network Map will help me to interact more with different group of people and hence stimulates me to think critically before giving my comments” and “the 21CC Profile graph as it is applicable in real life” were scarce.

5. Discussion and Conclusion

The qualitative findings around the benefits of self-referenced LA visualizations provide support for the quantitative findings reported earlier, particularly in terms of enhancing students’ critical reading fluency and engagement. Students’ comments about the LA visualizations contributing to increased self-awareness and engagement are in line with research on the associations between greater self-awareness and self-regulated learning strategies (Zimmerman & Martinez-Pons, 1988), emphasizing the potential of LA in making learning visible (Lockyer et al., 2013). Students’ emergent understanding of the influence of dispositional factors and social learning connections on learning intensifies the need for schools to give greater emphasis in developing these literacies and dispositions in students, given their influence on learning and life outcomes (Christakis & Fowler, 2009; Levin, 2012). The other shortcomings described by students will be taken into account in the design of future LA visualizations to provide more effective feedback.

Since the self-referenced LA student dashboard was randomly assigned to students within classes and across 7 classes, the benefits and shortcomings of using self-referenced LA dashboard visualizations reported in this paper are generalizable to some extent. However, we caution that the results should be interpreted within the context and sample – of one ability group in one Singapore secondary school. We acknowledge that the study can be further strengthened through the use of a quasi-experimental design with a control group as well as by evaluating the self-referenced LA visualizations against other types of LA visualizations. Nevertheless, though the findings presented here, we hope to provide some insight on the design and impact of LA dashboards in the Asian
educational context and contribute to better designed LA dashboards to maximize the learning potential of diverse students.

Acknowledgements

This paper refers to data and analysis from the research project NRF2013-EDU001-EL019, funded by the Singapore National Research Foundation, eduLab Research Program. The views expressed in this paper are the authors’ and do not necessarily represent the views of the National Institute of Education, Singapore.

References


Coloring Strategy Combined with Three-Dimensional Animation: Does the Mental Strategy Fits Everyone?

Jia-Yu CHU*, Yi-Ling HU*, Shi-Yu WANGb & Xiao-Qing GU*

*aDept. of Educational Information Technology, East China Normal University, China
bUniversity of Illinois at Urbana-Champaign, USA
*ylhu@deit.ecnu.edu.cn

Abstract: Spatial thinking is considered an important influence to learn Science, Technology, Engineering and Math (STEM) domains. The way to improve the spatial thinking ability drawn attention in the education. This study investigated a way in terms of multimedia environment combined with intervention on mental rotation strategy to train the spatial skill. 410 undergraduates recruited in the experiment revealed that the strategy intervention assisted with designed 3-D learning tool is a promising way to improve mental rotation ability of those who are not good at it.

Keywords: spatial thinking; multimedia learning; strategy intervention; 3-D animation and rotation

1. Introduction

Spatial thinking is a fundamental cognitive skill that has been recognized as a key ability to the performance in geography (Orion, Ben-Chaim & Kali, 1997), music (Hetland, 2000), visual arts (Goldsmith et al., 2013) and physical education (Pietsch & Jansen, 2012) as well as STEM disciplines (Newcombe, 2010). Given the importance of this ability, how to improve spatial thinking ability and evaluate the effectiveness of training has been a significant concern of educators. Some previous study shows that the spatial rotation in particular has been the object of training (Uttal et al., 2013), and even the spatial rotation strategy was found to be a key factor capable of improving the efficiency of mental rotation ability training. So what kind of strategy can be used in mental rotation testing? Does it fit everyone? This study focus on whether a mental rotation strategy which as an intervention in the multimedia environment combining with 3-D animation and manipulation could improve the mental rotation ability of young adults. Especially the divergence comparison of young adults between the United State and China.

2. Literature Review

According to numerous researches, the importance of spatial ability is beyond doubt, but the approach to enhance the ability still an unquenched aspect of educational practice. Some of them uncovered that computer games or augmented reality technology may enhance this ability. Lee (2016) combined augmented reality technology with 3-D holography in his experiment and found it is an effective way to reduce participants’ cognitive load and improve their sense of spatial direction compared to the traditional 2-D form.

The use of strategies is another way that influences the spatial ability learning. The researches (Gluck & Fitting, 2003; Linn & Petersen, 1985) adopted self-reported measures on exploring the strategies employed in the mental rotation. The research result indicated there were two strategies been utilized when people solving the mental rotation problems. They include: 1) Holistic strategy. A kind of strategy based on the rotation of the whole figure. 2) Analytical strategy. A kind of strategy that
mentally rotating a stimulus piece by piece or labeling parts of an object. Nonetheless, in general, both strategies will be involved during the mental rotation test.

Some research revealed that both spatial ability and spatial strategy were the factors affecting the effectiveness of spatial learning (Meneghetti et al., 2014; Stieff, Dixon, Ryu, Kumi, & Hegarty, 2014). Meanwhile, Geiser (2008) found that strategy intervention was useful to students who have good skills on mental rotation rather than those who have poor skills.

Therefore, this study aims to explore:
1) The effectiveness of the combination of strategy intervention with 3-D animation and manipulation.
2) Whether the level of mental rotation ability influences the effectiveness of the strategy intervention.

3. Methods

3.1 Participants

410 undergraduates (218 American and 192 Chinese) have enrolled in this experiment. The age of the participants varied between 17 and 23 (26.3% aged 17-18 years, 33.4% aged 19 years and 40.3% aged 20-23 years). During the experiment, participants were randomly divided into a control group (n=217) and an experimental group (n=193) to compare the differences of their test improvements after learning with different intervention strategies.

3.2 Materials

Questions required participants to finish is extended version of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (Revised PSTV: R). To make sure participants got enough training before post-test, researchers developed 10 new questions, the new questions are based on the shape of the original title with different rotation angles. The reliability of the extend version was evaluated by Cronbach alpha and the internal reliability coefficient is 0.905, which shows the question consistency in measurement.

In the pre- and post-test, images of a reference 3-D object before and after rotation is presented and then a testing 3-D object is presented with initial status and 5 images after different rotation. Participants are required to determine which of the five images is the correct one in terms of the equivalent rotation process as the reference object.

3.3 Strategy Intervention Tool

Between the pre- and post-test, a 3-D learning tool was provided to the participants, and specially, testing 3-D object presented to experimental group was with one face colored in the both initial states image and 5 rotated status images, the interface of learning process was showed in the right of figure 1. The colored face of the 3-D object could be a rotating stick to help participants imagine the image of rotated status. The intervention was designed to train participants’ strategies on mental rotation ability. Combined the strategy intervention with the 3-D learning tool, participants could manipulate and rotate the object manually to execute their thinking after known the answer of the question.
3.4 Procedure

Participants were required to take the experiment on the computer, for the technical can trace the process data. The experiment in both groups consisted of three phases:

- A pre-test with 10 questions without feedback of correct answers to assess the participants’ initial space rotation ability. Participants would get 1 point for each correct answer.
- Intervention phase consisted of two blocks. Each block consisted of 10 questions with feedback of question. Both control group and experimental group could play with the 3-D learning tool when the experimental group had strategy intervention.
- A post-test with 10 questions to determine, to what extent, the space rotation ability of both group has been improved.

4. Results

To examine the intervention effect of the 3-D learning tool to improve participants’ mental rotation ability and the influence caused by intervention strategies, the testing scores from the pre-test and the post-test, and a survey about attitude towards the experiment was collected. Data was analyzed with R, Microsoft Excel and IBM SPSS 22.

4.1 Learning outcomes

A paired-sample t-test was conducted to compare the scores of pre-test and post-test, Table 1 sets out the descriptive statistics for each group. Control group shows significant difference in the scores for pre-test(M=7.30, SD=1.81) and post-test(M=7.53, SD=1.79), t (216)= -1.99 p =0.048, experimental group have a better performance in post-test(M=7.25, SD=1.79) than pre-test(M=7.01, SD=2.10), but there is not significant difference, t (192)= -1.62, p=0.106. These results suggest that the 3-D learning tool without strategy intervention has significant intervention effect, participants in the control group made a better performance in test after intervention.

Table 1: Descriptive statistics of pre-test scores and post-test scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M±SD</td>
<td>M±SD</td>
</tr>
<tr>
<td>No strategy intervention (n=217)</td>
<td>7.30±1.81</td>
<td>7.53±1.79</td>
</tr>
<tr>
<td>Strategy intervention (n=193)</td>
<td>7.01±2.10</td>
<td>7.25±1.79</td>
</tr>
</tbody>
</table>
4.2 Effect of Strategy Intervention for Different Initial Rotation Ability Level

In the previous study, we found initial levels of the mental rotation ability have some influence to learners’ learning improvement. Therefore, factors could influence the performance included: 1) the strategy intervention provided by 3-D learning tool, 2) the initial levels of mental rotation ability. Therefore, next step is to examine whether the strategy intervention makes different influence for different level of initial ability. Participants were divided into three levels according to the pre-test scores. The mark of a low level is 60% accuracy, and the mark of a high level is 80% accuracy. The difference of post-test scores from pre-test scores was as an indicator to examine the improvement. Three one-way ANOVAs were conducted to compare the improvement of strategy intervention made at three levels. Table 2 sets out the descriptive statistics for each group in separate levels.

The result of one-way ANOVA showed a ladder by level increased. In level1 group, strategy intervention showed a significant effect of improvement, \( F(1,139) = 4.70, p = 0.032 < 0.05 \), Eta Squared \( \eta^2 = 0.033 \). In level2 group, there was no significant difference between control and experimental condition. In level3 group, participants in no strategy intervention condition showed better performance than those in strategy intervention strategy, \( F(1,120) = 5.92, p = 0.016 \), Squared \( \eta^2 = 0.047 \). The result indicates that strategy intervention has the opposite effect on groups with a low level of mental rotation ability and those with a high level.

Table 2: Descriptive statistics of improvement of different levels in both conditions.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 No strategy intervention</td>
<td>70</td>
<td>1.24</td>
<td>1.64</td>
</tr>
<tr>
<td>Level 1 Strategy intervention</td>
<td>71</td>
<td>1.87</td>
<td>1.81</td>
</tr>
<tr>
<td>Level 2 No strategy intervention</td>
<td>77</td>
<td>0.00</td>
<td>1.73</td>
</tr>
<tr>
<td>Level 2 Strategy intervention</td>
<td>70</td>
<td>-0.44</td>
<td>1.74</td>
</tr>
<tr>
<td>Level 3 No strategy intervention</td>
<td>70</td>
<td>-0.54</td>
<td>1.05</td>
</tr>
<tr>
<td>Level 3 Strategy intervention</td>
<td>52</td>
<td>-1.06</td>
<td>1.29</td>
</tr>
</tbody>
</table>

4.3 Effect of Strategy Intervention in Learning Process

Figure 2 presents mean scores in the four blocks, for each condition in three level groups. As can be seen in the chart, the group in strategy intervention condition performed a little behind the group in control condition. But during the learning phases, that is block2 and block3, group in experimental condition got higher scores, and specially, result of independent t-test shown significant difference in the score of block2 for no strategy intervention condition (\( M = 8.60, SD = 1.34 \)) and strategy intervention condition (\( M = 8.98, SD = 1.26 \)), \( t(406.73) = -3.0, p = 0.03 < 0.05 \).

![Figure 2](image-url)  
Figure 2. The mean scores of 4 blocks in control and experimental condition
Scores of group in strategy intervention condition fallen back to original state, which indicates that strategy intervention really provided a more effective strategy to deal with mental rotation questions, but this effect could not last after the support removed.

In order to understand how the strategy intervention influenced different level group, Figure 3 shows the mean of each block in a more detailed size.

Figure 3. The mean scores of 4 blocks in control and experimental condition

For level1 group, participants showed great improvement during and after intervention phases, level2 group only performed better during the learning phase, and for level3 group, the participants in experimental condition had initial level than those in control condition, but scored lower in the block3 and block4.

5. Discussion and Conclusion

The strategy is considered as a fundamental factor influence the performance of mental rotation ability and the effect of the train which has been verified in our study again. And further, we found that effective strategy support can increase the efficiency of mental rotation ability training. This study intends to compare the improvement of learning with or without strategy intervention. Though the tool involved in this experiment was examined in our previous research, we examined the effect of tool in both condition, and got the unexpected result that the group with strategy intervention showed no significant improve after learning when the control group had a great performance. Then the initial level of mental rotation ability has been used for the further analysis. The extraordinary findings shows that strategy intervention has a different effect to the participants with different initial levels. The members in the level 1 group may not be aware of using the strategy to solve this type questions, they would have developed the consciousness to use an effective strategy, and have been trained during the learning phase through 20 questions. Therefore, this group showed significant improvement in the following three blocks. Analysis of level 2 group showed no significant difference between two conditions, members in strategy intervention condition performed better than those in control condition until the strategy support was taken. They were divided into level 2 for scoring more than 6 points in the pre-test, which indicate they had ability to solve this type question. The intervention of tool may not provide a new strategy for them, but support their immature strategy as scaffolding. Therefore, those in experimental condition got a more effective support than those in control condition until the strategy support was taken. As for level 3 group, they got more than 8 points in the pre-test, which required talent or full-blown strategy to deal with the mental rotation. The strategy intervention seems like interference for them, it is hard to figure out whether strategy intervention influence their original strategy or not.

The result verified that there is not an effective strategy which could be used commonly. Aimed at students of different types or levels, different strategy and methods are needed. As for the learning tool designed by our team, it is suitable to be an assistant for the beginners or backward students rather than for those who have mastered techniques about mental rotation. In other hands, the learning tool may be beneficial to develop a normative mode of thinking to handle mental rotation questions.

In this study, we proposed a strategy intervention in the multimedia learning environment of combining 3-D animation and manipulation to support mental rotation ability development. The experimental study reveals that strategy intervention is a promising way to improve those with a low
level of mental rotation ability, but this way may not provide enough help for the group with a certain level ability. The results indicate that there is no such a mental strategy fit everyone; different learners need different support strategies.

6. Limitation and Future Work

This research has a few limitation. First, the question in the survey to investigate the strategy preferred by participants was not designed the well, the data could not provide reliable circumstantial evidence. Second, the amount of the question may be a little large, cognitive load may be an unobservable interference factor in the experiment. In the next version, we will optimize these parts of the learning tool, and future research might take account strategy in a more detailed aspect, to investigate its influence on the mental rotation ability improvement.

Acknowledgements

We would like to acknowledge Dr. Shiyu Wang from the University of Georgia State to help implement the study and collect the data.

References


Development of Alternative Conception Diagnostic System based on Item Response Theory in MOOCs

Yu-Cheng CHENG a*, Jian-Wei, TZENG b, Nen-Fu HUANG c, Chia-An LEE a & Meng-Lun Kuo d

a Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
b Center for Teaching and Learning Development, National Tsing Hua University, Hsinchu, Taiwan
c Netxtream Technologies Inc, Taiwan
d Institute of Service Science, National Tsing Hua University, Hsinchu, Taiwan

* s104062514@m104.nthu.edu.tw

Abstract: With the popularities of Massive Open Online Courses, a great number of enrollments in MOOCs generate much educational big data in terms of online activities and logs, which might be valuable for academia and practitioners. More personalized and intelligent online learning environment could be potentially created through educational data mining and learning analytics techniques. Based on Item Response Theory (IRT), the current study builds an item analysis system to identify alternative concepts/misconceptions from learners’ response in exams. By calculating difficulty parameter and discrimination parameter from massive learners, our systems are believed to benefit both teaching faculties and online learners. With the affordances of the system, teaching faculties could assess learners’ learning performance and quality of test items while alternative conception of learners would be identified for strategic learning. Other practical and technical implications will be discussed in this paper.

Keywords: Computer-Assisted Assessment, Item Response Theory, MOOCs (Massive Open Online Course), learning analytics, misconception, alternative conception, Item Analysis

1. Introduction

1.1 MOOCs

Massive online open courses (MOOCs) have emerged since 2012, providing high-quality online courses partnered with prestigious universities around the world. The innovative form of video-based learning has recently attracted attention from a variety of educational communities (Hsu, Huang, Chen, Lee, & Tseng, 2017). MOOCs allow learners from all over the world to enroll in a variety of online courses. Diverse learning modes and interaction could be found in MOOCs, including lecture videos, quizzes and exercises, discussion forums, learning visualization, etc. Rather than only providing traditional course materials like OpenCourseware, MOOCs are famous for lively interaction among stakeholders of the courses (students, professors, and teaching assistants). The overwhelming trend of MOOCs’ features makes it a contemporary way of 21-century learning (Huang, Hsu, Chen, Lee, Huang, Ou, Tzeng, 2017).

As MOOCs have been evolving, there have been two distinct types of MOOCs: cMOOC and xMOOC. cMOOC emphasizes the Connectivist view of learning, with a focus on learners’ autonomy and social nature of learning. xMOOC, on the other hand, is more closed to instructionism, implying the important role of lecture videos in the learning process (Breslow, Pritchard et al., 2013; Seaton et al., 2014). To be more specific, xMOOCs focus on the visual and audial representations in the instructional videos. Famous MOOC platforms such as Coursera, edX, and ShareCourse share such xMOOCs form of learning.

Jointly developed by National Tsing Hua University and Netxtream in Taiwan, SC (SC) is one of the largest MOOCs platforms in Chinese speaking society. More than 1,000 courses and over 100,000 active users have registered on SC since 2012. Furthermore, SC cooperates with over 70 universities and colleges. SC provides not only a good teaching and learning environment for online lecturers and learners, but also services like online discussion room, learning visualization system.
Although the developments of xMOOCs are widely accepted, the recent researches point out that the current progress of learning analytics systems is in a fancy stage (Ferguson & Clow, 2017). We believe that the great amount of educational big data in xMOOC platform might be valuable for course and platform improvement. IMS caliper and xAPI are both eager to accelerate the process of learning analysis through defined schema of learning activities (Santos et al., 2015 and Del et al., 2013). Among the types of activities, quizzes and exams are two primary ways of assessing learning performance in xMOOCs. Previous findings failed to focus on assessment mechanism. Moreover, most of analysis systems ignore the feedback of learners so that they lose the last arrow to close Learning Analytics Cycle or close the cycle in inefficiency way (Clow, 2012). There is still a lot of valuable information hidden in logs produced by xMOOCs platform.

1.2 Testing Analysis

Item response theory (IRT) is a statistical model used to evaluate the quality of measures (Kean & Reilly, 2014). Recently, there are several studies on Test-Item Analysis and Learning Management System(LMS). Fotaris, Mastoras, Mavridis & Manitsaris (2010) integrated LMS with an open source IRT tool, ICL, to improve the assessment quality, and found to (1) increase the efficiency of the testing process and (2) provide deeper guideline of test. Fotaris & Mastoras (2013) took advantage of IRT to optimize item pool on LMS. However, the author think the methodology is not limited to LMS. The environment of MOOCs with thousands of students might be more ideal for this experiment. Test-Item Analysis was also proposed to improve teaching and classification of students so that teachers could focus on learners’ weakness (Kiat,1981). It is a more stable and rigorous method, which will not be influenced by sampled students’ ability. The ideas of Item Difficulty and Item Discrimination are also thought of as practical way to estimate the item quality in traditional assessment.

However, the accuracy of IRT’s analysis is limited by scale of data. IRT has better performance on sufficient data. In order to solve this problem, Professor Deng first introduced Grey System Theory that could be applied on small amount data (Julong, 1989). Naggi also proposed Rasch Model Grey Student-Problem chart based on Grey Relation Analysis, one of the methods in Grey System Theory. Sheu, Tzeng, Tsai & Chen (2012) obtained misconceptions that cause learner’s confusion through the Rasch Model GSP chart even on insufficient data.

There have been some researches that perform IRT on data produced by MOOCs (Balint, Teodorescu, Colvin, Choi, & Pritchard, 2014 and Balint, Teodorescu, Colvin, Choi & Pritchard, 2015). To our knowledge, the current MOOCs services do not equip with item analysis features. This might be due to two reasons. First, the diverse and considerable number of learners deter the development of dynamic assessment system. Computing timely and precise data requires expansive computational power. Second, platform owners and designers put more emphasis on the interface and interaction design of the MOOCs platforms, failing to take advantages of assessment mechanism to improve learning effectiveness. Assessment-based learning is believed to be effective and goal-driven way of online learning. Through effective design of assessment system, learners’ motivation and self-efficacy might be increased to foster learning. Therefore, this study developed a service integrating Test-analysis System into an existing MOOCs platform to offer alternative conception diagnostic service and testing item analysis service for both teaching faculties and online learners.

2. System Architecture

The whole service is built on two running instances, one is MOOCs service instance and the other one is datacenter. MOOCs service instance provides common online course as regular MOOCs platform like edX and Coursera. Datacenter instance provides restful API for the communication between two instances. MOOCs instance could send user’s activity logs to data center and request for advanced analytics. This architecture provides flexibility for datacenter providing service to other MOOCs platforms and independency of two instances.

2.1 MOOCs Platform
In this paper, the MOOCs platform is SC. To communicate with the datacenter, the task of SC is required to sends logs with de-identification ID to datacenter through HTTP request. When students perform some activities on course material, activities logs will be packaged in JSON format, and send to the corresponding API route with specified header fields. The raw data will be stored in database waiting for analysis. After processing, the results will be also available through corresponding API route. SC sends a HTTP request and reveals advanced analytics to teachers and students. The visualization and results tables are also available on our service.

2.2 API Server

The API server is the only interface that datacenter could connect to outside by. This would avoid exposing database and internal structure to external. It would validate the activity log received, filter logs that are from unknown source, and store the logs in specified schema in the database. Note that all the records stored on server are processed by de-identification. MOOCs doesn’t send records with user’ privacy, such as school, age, country, etc. Only unique identification is used to distinguish whom the records belong to. While MOOCs instance requests for certain analytic results, the API server fetches the results from database and returns the results to MOOCs instance similarly.

2.3 Alternative Conception Diagnosis Module

This module retrieves the misconceptions based on ordinal of gamma value, a float number list ranging 0 to 1. The problems with gamma value near 0.5 are considered as misconceptions. The alternative conception points out the concept that students need to pay more attention to. Moreover, it is proved that results can be produced under of even a small amount of people such as classical face-to-face education (Sheu, Tzeng, Tsai & Chen, 2012). This overcomes small number of passed students problem caused by the low pass rate problem while performing other analysis.

2.4 Analyzer

The analyzer is triggered by crontab daily. It calculates Testing Items, such as Index of difficulty and Index of discrimination for each exercise. Besides, it also mixes the result of Alternative Conception Identify Module with problem-concept mapping table to extract alternative conception of each course. All the analytics results above will be stored back to database.

![Diagram](image)

**Figure 1.** An insight view of Datacenter including four modules, Database, API Server, Misconception Diagnosis Module and Analyzer.

3. Algorithm
In this section, we illustrate the algorithm which analyzes answer records stored in database and comes out with Item difficulty index, Item discrimination index and the alternative conception questions.

### 3.1 Testing Item Index

In this paper, Kiat’s definitions of high-low group Difficulty Index and Discrimination Index are applied. In the beginning, the program loads the learner’s records of one course from database. Because of the high attrition rate and low completion of course content in MOOCs, the records of these learners who didn’t complete the course would bias the analysis results. The following process includes only the records of learners who completed the course and took all the exams.

Before calculating, it is necessary to label student by their grade. If the student scores in top 27%, the student is labeled as ‘high score group’. On the contrary, the students in last 27% are labeled as ‘low grade group’.

The analyzer will aggregate the records of each problem and group by above-mentioned labels. Finally, it could combine the correct rate of the two groups to figure out the Item Difficulty Index and Item Discrimination Index. The difficulty index is within the range $[0, 1]$ and the discrimination index is within the range $[-1, 1]$. The indexes of problems are all stored back to the database eventually.

Because it’s not friendly to show raw index value to users and learners and tutors would not understand the meaning behind the value, the raw index value is transformed into signals like traffic light, great, normal and bad, so that users could easily catch the hint. These signals are shown in Chinese text on SC.

### 3.2 Alternative conception

In this paper, the methodology proposed by Sheu, Tzeng, Tsai & Chen (2012) is used to calculate the grey relation and to retrieve the alternative conception of the course. Unlike other algorithm, the gamma value is between $[-1, 1]$, which make it easier for comparison. Nagai’s equation used in aforementioned paper is shown as follows.

$$
\Gamma_{0i} = \Gamma(x_0, x_i) = \frac{\overline{\Delta}_\text{max} - \overline{\Delta}_{0i}}{\Delta_{\text{max}} - \Delta_{\text{min}}}
$$

where $x_0$ is partial grey relation’s reference vector and $x_i$ is comparative vector and

$$
\overline{\Delta}_{0i} = \|x_{0i}\|_\rho = \left(\sum_{k=1}^{m} |\Delta_{0i}(k)|^\rho\right)^{\frac{1}{\rho}}
$$

Where $\Delta_{\text{max}}$ and $\Delta_{\text{min}}$ represent $\overline{\Delta}_{0i}$’s maximum and minimum. When $\Gamma_{0i}$ is close to 1, it means that $x_0$ and $x_i$ are highly related to each other. On the other hand, if $\Gamma_{0i}$ is close to 0, the relationship between $x_0$ and $x_i$ is lower.

### 4. Experiments and Results

In the experiments, two courses on SC are selected: Investment and Introduction of computer network. Both course have been taught so that there are students’ responses of exam collected. We applied analysis on these records and provide results to processing course on the platform. In the course of Investment, our analysis focuses on 240 participants who completed both of the exams and our analysis in Introduction of computer focuses on 372 participants who completed both of the exams too. We also have requested for professor’s agreement and asked teaching assistants for problem-concept mapping tables.

Figure 2 shows that students could see the discrimination index and difficulty index after the exam assessed. Note that we only show Chinese signals to represent level of discrimination index and difficulty index. Besides, we also provide these indexes with weekly exercises and item pool of tutors. Report of index reveals the quality of whole exam is provided to tutors, helping tutors to understand problems that need to be improved. At last, we show list of alternative conceptions in scope of midterm, as shown in table 1, to guide students to review the course material about these alternative conceptions.
Both Discrimination parameter and Difficulty parameter are attached to each exam problem.

Table 1: An example of misconceptions list of Investment midterm

<table>
<thead>
<tr>
<th>Name of Course</th>
<th>Misconceptions of Midterm (In Chinese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment</td>
<td>金融市場,資本市場,股票指數,融券放空</td>
</tr>
</tbody>
</table>

5. Conclusions & Future work

Due to the rapid development of MOOCs and the great number of online learners, massive online activity logs data would be generated and stored in web service. Online learning environment is different from traditional face-to-face classroom interaction, which teachers would be easier to understand learners’ feedbacks for improving teaching quality. Moreover, the massive enrollment makes it even challenging for online instructors and teaching assistants to provide with adaptive intervention and guidance.

With the affordances of learning analytics, this paper develops an Alternative Conception Diagnostic System, automatically providing item analysis for teachers and alternative conceptions for online learners. Moreover, the system cooperates with a MOOCs platform that serves thousands of learners. This intervention makes learners identify alternative conceptions of course so that they could devote more efforts on these sections to get a higher performance. It also closes the learning analytics cycle and makes the online learning environments a better place.

Future work is to improve the presentation of alternative conceptions and figure out further usages. In this paper, only text table of misconceptions is produced. We are seeking for a better representation of alternative conceptions, such as knowledge map, which shows organized structure of course concepts. Highlighting the alternative conceptions on the Knowledge Map is a straightforward way to make learners understand, promoting the learning efficiency of MOOCs.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document.

References


Capturing Changes and Variations from Teachers’ Time Series Usage Data

Longwei ZHENG*, Yuanyuan FENGb, Rui SHIa & Xiaoqing GUa

aFaculty of Education, East China Normal University, China
bSchool of Foreign Languages, East China Normal University, China

*lwzheng@dec.ecnu.edu.cn

Abstract: The type of time series data is very common in educational settings, such data set usually provide usage trajectories involved complex contexts and unknown noisy. In this paper, we describe a statistical analysis method to detect a switch point over teacher’s comment data set which is collected from a student evaluation system. We find that it is indeed possible to employ a statistical model and Bayesian inference to detect teacher who increased, decreased or no-changed their behavior. We also illustrated the potential of the application of change detection to conduct data exploration, which can be relative to understand teacher technology adoption and usage transition.

Keywords: change detection, teacher’s technology adoption, statistical modeling, MCMC

1. Introduction

In educational settings, time series data is often applied for possible dropout, outcome predicting, learning pathway discovering, and engagement detecting (Arnold & Pistilli, 2012; Blikstein, 2011; Clow, 2013; Park et al., 2017). With this type of data set, statistical change detection techniques can be utilized to investigate segmentation or human behaviors, such as detecting special change time on individual's clickstream which is a widely studied topic in web interacting research (Carlin, 1992; Park et al., 2017). As in most cases of change detection study, the main contribution of change detection is to provide a feature to group users into similar transition groups or search a signal to partition a time series into two distinct observations (Eckley, 2011).

The purpose of this study is to investigate teacher’s time series usage of a student evaluating system by employing a simple change detection technique which is introduced by Davidson-Pilon (2015). The comments which created in a student evaluating system will be computationally detected significant switch point of number of comments. We hope to identify teachers who increase, decrease or show no change in comment dataset, so that we can get more insights of teacher’s technology adoption and usage by discovering the relations to other aspects of technology usage between before switch point and after.

2. Methodology

2.1 Data

To explore this issue, we collected teachers’ comment data from a student evaluating system which supports teacher to code the students’ performance by clicking on a tablet application, and then records the code data into system to automatically generate a weekly review for each student. The users of this platform are teachers at 8 primary schools of Shanghai. We obtained 991,979 comment records for 2,224 students in 59 classes which has almost same class size over a 4-month semester. Each comment records the praise or criticism that a teacher evaluates a student at a certain time. Typically, a teacher would click a button that might indicate ‘You have all the right answers in arithmetic quizzes today! Good job!’ if he or she wants to commend a student’s performance in class. As the teachers’ contexts (such as grade, subject, students’ contexts, etc.) are various, the comment usage and comment content can vary from one group of teachers to another.
2.2 Modeling of Switch Point

A Poisson random variable is appropriate model for the type of count data of teacher comment records. We assumed that time is a discrete with $T$ discrete days and $t = 1, 2, ..., T$ being as index running from the first day to the last day of period in which an individual teacher $i$ is active in the system. Therefore, day $i$’s comment count can be denoted by $C_{it}$,

$$C_{it} \sim \text{Poisson}(\lambda_i)$$

(1)

The expected value of the Poisson distribution is equal to its parameter $\lambda_i$, we assume that the $\lambda_i$ on day $\tau_i$ which is independent from a uniform distribution, the parameter $\lambda_i$ suddenly jumps to a higher value or declines rapidly to a lower value. Such a sudden transition can be denoted as two $\lambda_i$ parameters, one is assigned for the period before $\tau_i$, and the other one is assigned for the rest period. In addition, the exponential distribution is employed to model $\lambda_i$, so the parameter of exponential distribution $\alpha_i$, which reflects the prior belief for $\lambda_i$, and is included in the switch point model:

$$\begin{align*}
\lambda_{i1} & \sim \text{Exp}(\alpha_i) \quad \text{if } t < \tau_i \\
\lambda_{i2} & \sim \text{Exp}(\alpha_i) \quad \text{if } t \geq \tau_i
\end{align*}$$

(2)

Thus the expected value of the comment count can be identified as:

$$\frac{1}{N} \sum_{t=0}^{N} C_{it} \approx E[\lambda_i | \alpha_i] = \frac{1}{\alpha_i}$$

(3)

Figure 1 shows a teacher’s daily comment count data in the completed observing period. We can find a significant decrease from early half period to latter half period, so we assume that this teacher changed his or her behavior on day $\tau_i$ by switching $\lambda_{i1}$, the parameter of comment count $C_{it}$, to a higher value $\lambda_{i2}$.

2.3 Bayesian Inference and Markov Chain Monte Carlo

In order to estimate parameters of switch point model, we employed Bayesian inference, that is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available (Box & Tiao, 2011). Markov chain Monte Carlo (MCMC) is a general method for simulation of stochastic processes having probability densities known up to constant of proportionality. MCMC is often applied to Bayesian inference that measure and update the believability of events after considering new evidence (Green, 1995). The algorithm can be expressed as an approximation solution to the posterior: the way that the algorithm moves in the general direction towards the regions where posterior distribution exist, and collect samples based on the the position’s adherence to the data and prior distributions. Once it reaches the posterior distribution, the samples can be collected as they likely belong to the posterior distribution (Davidson-Pilon, 2015).

In this study, we utilized Bayesian inference to get the posterior distributions of three parameters ($\lambda_1, \lambda_2, \tau$) of change detection model by performing MCMC. Then we get the expectation from each parameter distribution and assign to every teacher. As shown in Figure 1, a teacher is detected that he or she has an obviously higher count of comments after day 33, and two $\lambda$s of two distinct periods are also estimated with their expectations.
3. Result

3.1 Change Detection

We omit the data of teachers who has a very limited usage (count of active days < 20). We also restricted the switch day to be ranged from the 10th day from top to the 10th day from bottom, since switch point detection at the beginning or end of tends to be unreliable due to small sample sizes or be affected by semester timelines, and also not so meaningful in terms of interpreting (Park et al., 2017). Therefore, the usage which is detected switched at this two period will be considered as no-change.

Finally, we applied the change detection methodology to comment data performed by 145 teachers. The numbers of teachers detected as belonging to increase group and decrease group are shown in Table 1. The result shows that more teachers (42.07%) increased comments in the observing period. There is also a significant difference between two group’s distributions. If we order the teachers by expected number of the first four weeks ($\lambda$), we can find that the teacher who has intensive usage has more possibility to increase comment rate, and the teacher who has more moderate or limited usage tend to decrease comment rate. Therefore, this change detection result, to a certain degree, can be considered as a feature when predicting teacher’s technology adoption in early days.

Table 1: Change detection result summary.

<table>
<thead>
<tr>
<th></th>
<th>Number of Increased Teacher</th>
<th>Number of Decreased Teacher</th>
<th>Number of No-change Teacher</th>
<th>Number of Analyzed Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>61</td>
<td>46</td>
<td>38</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>42.07%</td>
<td>31.72%</td>
<td>26.21%</td>
<td></td>
</tr>
<tr>
<td>Top Half by ordering $\lambda$</td>
<td>41</td>
<td>67.21%</td>
<td>15</td>
<td>73</td>
</tr>
<tr>
<td>Bottom Half by ordering $\lambda$</td>
<td>20</td>
<td>32.79%</td>
<td>31</td>
<td>21</td>
</tr>
</tbody>
</table>

Figure 2 provides a day-by-day summary for the number of detected teacher changes per day. There is an obvious pattern in this visualization. The bar plot shows that there are many changes detected (more than half teachers from increased group increased their comments) from 45th day to 60th day, which is the followed two weeks after midterm. This, to a certain degree, can agree with the intuition that some teachers adopt this system to help them evaluate student after a half-semester probationary using period.
3.2 Discourse Variation

In this section, we applied switch point detection result to partition time series to two distinct data sets. Discourse variation analysis will be conducted by comparing the data set before switch point and the after.

Firstly, we utilized Shannon Entropy to calculate variety of the comments that was given by teacher. Shannon entropy is a mathematical construct to model variety from information theory (Shannon, 2001).

As shown in Table 2, there is significant difference of comment variety transition between two switched point detected groups, it makes sense that teachers from increased group have more chance to give more comment for more students. However, if we partition each group to two sub-groups by median number of types of comments that are given by every teacher ($N_{ct}$, the median of population is 11), we can find two distinct different results. Again it makes sense that the teachers who are from increased group and prefer to use specific words have more probability to increase variety (the percentage of the Entropy increased is 66.67%). We can also see the teachers who are from comment decreased group and prefer diverse comments have more probability to increase entropy (the percentage of the Entropy increased is 70.00%). It indicates that these teachers might tend to make specific comments to specific students and avoid to give high-coverage comments after their switch points, such behavior change will lead to a comment decrease detection.

Table 2: The change summary of comments variety.

<table>
<thead>
<tr>
<th></th>
<th>All Switch Point Detected</th>
<th>$N_{ct} \leq 11$</th>
<th>$N_{ct} &gt; 11$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Entropy Increased</td>
<td>Entropy Decreased</td>
</tr>
<tr>
<td>Comment Increased Group</td>
<td>61</td>
<td>37</td>
<td>60.66%</td>
</tr>
<tr>
<td>Comment Decreased Group</td>
<td>46</td>
<td>22</td>
<td>47.83%</td>
</tr>
</tbody>
</table>

Secondly, we compared the sentiment transition between two groups by measuring the comment’s sentiment. The sentiment measuring process consists of following steps: 1. Segment each comment to a set of words tagged with part-of-speech (such as noun, name, verb and etc.); 2. Build sentiment corpuses by labeling positive or negative for segmentation result; 3. Assign degree weights
to adverbs by employing a dictionary of degree of Chinese adverbs (Dong & Dong, 1999), such as the superlative word (‘most’, ‘best’, ‘extremely’), comparative words (‘more’, ‘a bit’, ‘much’), negative words (‘not’, ‘no’, ‘but’); 4. Compute the sentiment score of each comment by multiply degree weight and set sign according to its sentiment corpus.

The sentiment analysis result shows a significant difference of quantity between the positive comments (958,442) and the negative comments (33,537). It illustrates that teacher are inclined to make more encouraging evaluations or positive praises to students. But we also find that most teachers began to adopt more negative comments over time. Thus we select the teachers who has a sentiment score difference which was statistically significant between two periods.

As shown in Table3, more teachers decreased sentiment significantly after switch point, and there is no significant difference between the comment increased group and comment decreased group. However, if each group is partitioned to two sub-groups by median number of types of comments that are given by every teacher, again we can find that the teachers who prefer diverse comments from either group have more probability to increase comment sentiment. As mentioned in entropy analysis above, this is to be expected with less negative comment behavior since these teachers’ decrease for less-frequently-used comment in latter period. On the other hand, the teachers who prefer same comments are more likely to adopt more negative words after switch point.

Table 3: The change summary of comment sentiment

<table>
<thead>
<tr>
<th>All Switch Point Detected</th>
<th>$N_{ct} \leq 11$</th>
<th>$N_{ct} &gt; 11$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Sentiment Increased</td>
</tr>
<tr>
<td>Comment Increased Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>23</td>
<td>43.06%</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>69.23%</td>
</tr>
<tr>
<td>Comment Decreased Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>17</td>
<td>51.52%</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>66.66%</td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>48.78%</td>
</tr>
</tbody>
</table>

4. Discussion

This paper introduced an approach for change detections based on teachers’ comment records from a student evaluation system. It is indeed possible to employ a statistical switch point model and Bayesian inference to detect teacher who increased or decreased their action rate, even the count data of user action generally pose its noisy nature. We also illustrated the potential of the application of change detection to distill features, which can be relative to teacher technology adoption. In our study, this approach to exploring teacher technology use behaviors from searching transition signals have been very revealing. They help us to understand that the teacher who has intensive usage in early days has more possibility to adopt technology in future. Further, modeling technology use in terms of detected change result combined with other dimensions, such as variety and sentiment, may be a useful diffuse framework for capturing the usage pattern variation and characterizing how usage transition happens. In our study, the feature of comments diversity may be helpful to explore why the changes of variety and sentiment appear in different groups. In the future, our work involves the discoveries of more unknown behaviors within technology usage of teacher by multi change point detections.
References


TGlass: A Custom-made Wearable Promoting Accessibility for Tetraplegic

Maikon SOARES*, Felipe BEDÊb, Thiago ARAUJOa, Daniel CAVALCANTEc, Adriano FREITASb, Zuila CARVALHOc & Francisco OLIVEIRAb

*aCeará State University, Brazil
bUniversity of Fortaleza, Brazil
cFederal Institute of Ceará, Brazil
*maikonigor@gmail.com

Abstract: Although the law of quotas enables people with disabilities (PwD) to join the labour market, better paying working positions require more skilled professionals, thus demanding specific professional qualification. Tetraplegic people often struggle to access regular courses, such that online courses could be an alternative to offer them some technical formation. However, they demand special equipment to interact with online courses. In this context, this paper presents the TGlass, a low budget smartglass designed to be distributed as companion material to online courses, respecting users anthropometry and providing them with comfort and movement freedom. Its input/output peripherals together with the eye tracking software technique makes possible the communication between PwD and a learning environment. As a result, we expect that a large scale production will reach a quite cheap price for the device, which will be able to better the life of quadriplegic people as for their professionals opportunities and thus to their quality of life.

Keywords: TGlass; Quadriplegic; Distance education; Accessibility; People with Disability.

1. Introduction

The number of places in the field of Information Technology and Communication (ITC) has been continuously growing, over 40% in 2015 and 30% in 2016 (Zogbi, 2016). Such field offers great opportunities for programmers, some of which are destined for people with disabilities (PwD). In Brazil, there are about 45 million of PwD (Census, 2010), who commonly struggle to access technological resources that could facilitate their learning process towards a better professional qualification, especially those who are tetraplegic or quadriplegic - those with no voluntary movements of the upper and lower limbs. Soares (2014) highlights the need for technological solutions that provide PwD with a better qualification so they can overcome their difficulties. We carried out a wide investigation on the related work addressing this problem in the past ten years, though only a few were found reporting solutions on that matter. These are briefly summarized next.

Guerreiro (2008) proposes a mobile system called “The Myographic Mobile Accessibility for Tetraplegic”, which allows quadriplegic people to use the smartphone by sending commands through muscle stimulus. The system consists of four modules: 1) user interface (UI); 2) information presentation (IP); 3) electromyographic control (EC) and 4) application. The UI module communicates with the application module sending and receiving data by the EC and IP modules, respectively. The user wears a type of vest containing electrodes that pick up the muscular stimuli, processes them and sends commands to the system. The application module runs in on Windows operating system. Bluetooth technology is responsible for maintaining connectivity between the modules, serving as a communication channel. Finally, the IP module is responsible for returning the system responses to the user. The information depends on the state of the application. The messages can be displayed via audio or visually.
The Myographic Mobile solves the personal computer data input for quadriplegic problem, however we believe that our approach, in opposition to the proposed vest, brings more benefits to the users because they will have more control over the device that will capture, in a less susceptible to noise, their intentions. In this investigation we aim to take advantage of the natural movements of the human body, to catch data input from user, in a comfortable way.

Caltenco (2010) propose the Tongue Wise system. A software developed to interact with a computer through a set of sensors located at the tip of the user’s tongue. The tongue-controlled system consists in a wearable device that contains 18 sensors capable of covering most standard keyboard and mouse commands. The sensors were subdivided into 2 types: 10 for keyboard commands and 8 for multi-directional commands that simulate the movements and clicks of a mouse. Each movement picked up by the sensors is converted into the standard keyboard command, called the key code, or mouse movement pattern. The generated commands are sent via USB port to the computer generating a visual display on the operating system. The usability tests report that the distance and location of the sensors directly influence the speed and precision of use, and conclude that the best location for the sensors is on the tongue and not on the sides of the cheek or roof of the mouth. The typing accuracy was about 2.8 correct words per minute (CWPM) with a standard keyboard and 1.7 CWPM with Tongue Wise system.

The eye tracking system proposed by Ye (2012) aims to detect moments of eye contacts between an adult and a child. His method utilizes commercial smartglass that provide eye gaze tracking technology to determine the adult gaze point and combine this with video analyzes from the child’s face. The device is similar in appearance to regular glasses, with an outward looking camera that captures a video of the scene in front of the examiner.

Vidal (2014) presents an interface design for wearables with near-eye display. These devices present the user with a duality of visual worlds, with a virtual window of information overlaid onto the physical world. The author suggests that the wearable interface had to benefit from understanding where the user’s visual attention is directed. The eye tracking technique is used to create and analyze interfaces, it can notice when user’s gaze is into the application or real world, outside application boundaries. The author takes into consideration some aspects such as when the user looks at an extreme edge of the screen or a display menu inside application. On one hand, research has shown the plausibility of integrating gaze into near-eye displays.

In this work, we propose the Tglass, a low budget smartglass designed to be distributed as companion material to online courses.

2. Methodology

2.1 Virtual Learning Environment (VLE)

The access of PwD to regular schools is still poor. Knowing that, we provide an accessible E-learning system where PwD can do courses and develop some skills. All the material and tools are developed by an interdisciplinary team made of programmers, pedagogues, language signers, tutors and other professionals and then tested and validated by a team of PwD before being published in the platform, such that all the content is already accessible for people with hearing or visual disability and is currently being adapted to quadriplegic people.

Quadriplegic users of computers report that handling a device in their mouth for interacting with the machine is completely tiring and uncomfortable. Therefore, we believe that a proper device to these people would work based on voice commands or on small movements of the head or of the eyes, besides being light and comfortable so they can do the activities without much effort.

2.2 The Device usability

In this work, we propose a low-cost device named TGlass, which consists of a pair of glasses with sensors that detect head movements and a software that operates upon each command.
TGlass is designed for tetraplegic users over 16 years of age whose level of the cervical lesion is below C5. Tetraplegic will need someone to adjust the glasses to their head to make them firm and ensure that the input and output peripherals are correctly capturing the user’s movements.

With TGlass, the user will be able to send commands to the computer by using their own voice or by moving their eyes. The device will enable the capture of intentions through ocular manifestations. Speech is captured by the microphone and then submitted to natural language algorithms in order to process the user’s intentions and act upon the VLE.

TGlass will be shipped pre-configured with the student’s login and password so that the user can initiate activities immediately after dressing and adjusting the equipment. After initialization, the initial menu containing information related to the course they are enrolled will be displayed on the TGlass screen, and the option to choose a learning object from the lesson, exercise, support material or evaluation. The user chooses the option by pointing with his/her eye the desired icon, and with the blink of the eyes selects the chosen option.

Initially the field of view, merged into a virtual minimalist interface, is projected into the user’s eyes. Then, following the eye movement, the user navigates through the existing options and, with a blink of an eye, selects the desired activity. At the beginning of the activity, the application, using augmented reality, plots the first step of the object model over a predefined target on the user’s workspace. Through manual gestures, the user can optionally request help to find some of the parts needed to assemble the object. When you complete the step, the system makes the recognition of the mounted object and validates the assembly, releasing the next phase in the event of success.

The design of the interaction considers the Fitts’ Law, an empirical model that analyzes the accuracy and speed to perform an activity. According to Fitts (1954), the time required to move to a target area is a function of the ratio of the distance to the target and the width of the target. Such rule originally applies to only one dimension, but there are extensions for two-dimensional activities with direct applications in computer systems in activities based primarily on point and click. In this context, it stems from Steering Law (Accot & Zhai, 1999). Therefore buttons or widgets with a larger size are used in order to reduce the time spent and the difficulty to reach them.

Among the available features to facilitate navigation we can mention:
- Move the mouse cursor by detecting the movement of the iris through the camera.
- Simulation of the mouse click when the user flashes the eye or by resting a certain time limit on the desired target.
- Larger targets in the VLE interface to facilitate navigation and selection of available functions.
- Attracting targets can be activated to attract the cursor if it is in a proximal region.
- Voice commands to access specific features.
- Writing texts via voice recognition.

2.3 Device Assembly

In order to develop the device, we have taken into consideration some important aspects regarding its design and form factor, such as weight and robustness, respecting the anthropometry, cost, comfort and movement freedom.

For building this wearable device, we had to consider the physical particularities of someone’s head, by means of parameters such as the head format, nasal bridge format, distance between the eyes and general facial geometry dictate the equipment measures. Alves (2011) found that the average Brazilians’ head dimensions were: head’s circumference (horizontal perimeter): 51.93cm in men and 53.56cm in women; head width (frontal diameter): 15.2cm in men and 15.65cm in women; and head length (profile diameter): 19.12cm in men and 19.75cm in women.

The glasses will be produced in a unique size, but soft adjustments will be implemented on the glasses, such as size regulation in the lateral temple and also width regulation. The whole glasses frame measures 60cm of width and 30cm of length, so that both men and women can wear them.

A sketch of the physical structure of the TGlass frame is presented in a variety of angles in Figure 1. The front part is equipped with lenses, a microphone and a video camera. These accessories compose the peripheral input and output of the glasses, so that the user can access the computer.
commands and receive the feedbacks of the system in use. The left and right temples are flexibly fixed with a length adjustment that fits to the users’ ears, which makes sure that the glasses will be fixed to the face. The upper part is made with a rim where the width adjustment flaps and the peripheral input and output are.

![Figure 1. Physical structure of the TGlass.](image)

The weight and duration of the battery are also crucial factors that should be taken into consideration when projecting this sort of device. The average time that a user spends to accomplish the activities in VLE is 60min for deaf people, reaching 80min for people with motor disabilities. Knowing that, the TGlass does not need powerful batteries as tablets and the notebooks do. A smaller battery reduces production cost as well as the glasses weight, thus assisting the user with comfort and mobility.

The TGlass has in its physical structure a series of components that enable data input/output for communicating with the VLE. Front and back video cameras provide the recognition of the working area and visual tracking. An LCD screen shows the user the VLE feedbacks, and a microphone enables the use of voice as input. These peripheral details allow the user to interact with the VLE content without their hands.

Figure 2 shows how the peripheral components are disposed in the physical structure of the TGlass. In Figure 2-A, we see the TGlass front camera, used for scanning the workspace in augmented reality activities, and user feedback is given through an LCD display. Figure 2-B depicts the back camera of the glasses, which will be pointed out to users’ eyes to capture images of eyes reactions; from the generated images, users will be able to reach the clickable targets of the VLE with just one look, once a blink of an eye works as a mouse click. In Figure 2-C, we show the micromirror DLP2010, from Texas, responsible for the projection of the virtual environment that assists on the augmented reality tasks. This component has a resolution of 800x600px and it shows the images using 16bits of colors, which is enough to enable the VLE proposed activities. In Figure 2-D, we present the lenses for projections of LVE content in the users eye. These lenses have a semi-transparent mirror that allows the passing of light, making it possible for the user to see both the external and virtual surroundings, thus providing a better immersion on the virtual environment. Figure 2-E shows the microphone of the device, which enables voice commands, where texts in blocks and navigation commands via voice can be entered by the user, being free from the hands to write something.

The base structure, or the glasses frame, are produced by a 3D printer, which makes its confection cost cheaper and accelerates the building of the parts for reposition. In order to guarantee low weight and comfort, the material chosen for printing is the PLA 1.75 mm, a light but resistant and easy to find material.
In Table 1 we list the all the components used for building the TGlass and its prices.

Table 1: Cost of the components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Cost (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raspberry Pi A+</td>
<td>23.00</td>
</tr>
<tr>
<td>Camera Module Raspberry</td>
<td>8.62</td>
</tr>
<tr>
<td>USB microphone</td>
<td>7.90</td>
</tr>
<tr>
<td>USB Micro Camera</td>
<td>4.15</td>
</tr>
<tr>
<td>Powerbank 2000 MAH</td>
<td>5.00</td>
</tr>
<tr>
<td>3D printing</td>
<td>2.50</td>
</tr>
<tr>
<td>DLP2010 (LCD)</td>
<td>2.50</td>
</tr>
<tr>
<td>DLPDLCR2010EVM-PCB (DRIVER)</td>
<td>249.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>About 374.45 US</strong></td>
</tr>
</tbody>
</table>

3. Final Considerations

In this work, we propose TGlass, a device that provides accessibility to a VLE for quadriplegic so that they can have access to professional courses without much effort. The target audience are those with a cervical injury level below C5, since the cost-benefit ratio of people with high (C1-C4) injury would not be favourable. PwDs that fit the profile of the target audience need good speech skills so that they will be able to perform effectively in our platform.

The elevated initial price is due to the production of only one single Tglass. However, when taking a greater production into consideration, cost considerably decreases. We expect this device will have a low cost of construction and acquisition compared to the costs of personal computers and accessibility equipment in the market.

Such device represents an innovation in the field of distance education, especially when focusing on people with disabilities in the context of quadriplegics. Additionally, we state that it will provide easy handling and configuration for tetraplegic and for anyone else.

We believe TGlass will bring great benefits for PWD, once it will favour and enable greater interaction of these people with the outside world, thus reducing the visible and invisible barriers that quadriplegic people experience, becoming a tool of social inclusion and online education.

References


Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6, 381.


Augmented Reality based Learning Support System for Mental Rotation

Midori NAKANO*, Yukihiro MATSUBARA, Masaru OKAMOTO & Noriyuki IWANE
Graduate School of Information Sciences, Hiroshima City University, Japan
*lnakano@lake.info.hiroshima-cu.ac.jp

Abstract: In the field of figure, it is said that learners’ subjective “manipulation” enhances the learners’ interest in mathematics and improves the degree of learning. In addition, spatial cognitive and mental rotation abilities can be considered necessary to solve the questions on spatial figures. This research constructs a system that enables learners to manipulate marker and learn mental rotation questions. The proposed system realizes learners’ subjective manipulation by freely manipulating stereoscopic markers. This system makes it possible to confirm spatial figures from various angles by using stereoscopic Augmented Reality markers.

Keywords: Mental rotation, Augmented Reality, spatial figures

1. Introduction

Spatial cognitive and mental rotation abilities can be considered fundamental to solve the questions on spatial figures. Mental rotation ability can be acquired by learning using stereoscopic objects. It is said that subjective “manipulation” by learners enhances learners’ interest in mathematics and improves the degree of learning (Yuki, Shota and Tokuji, 2013). In many cases, learning is performed using paper media such as textbooks, while learning via a three-dimensional object is not performed very often. In order to address this issue, Kei and Nobuyuki (2014) conducted learning support on the mental rotation questions. The system they propose can virtually observe three-dimensional figures using Augmented Reality (we will abbreviate it as AR hereafter). However, since Kei and Nobuyuki’s system recognizes the AR marker affixed to the desk, it seems difficult for the learner to move hands on a subjective manipulation.

This research constructs a system that enables learners to learn the space figure by operating the marker themselves. By using stereoscopic AR markers, this system realizes the learners’ subjective manipulation, making it possible to see figures from various angles. In the proposed system, a state in which a learner manipulates two types of markers is recorded by a smartphone; then, the CG of the three-dimensional object is superimposed on the image and displayed on the smartphone screen.

2. Proposed System

The proposed system is constructed using a smartphone and a marker. The right side of Figure 1 shows the appearance of the proposed system. The left side of Figure 1 shows the screen of the system. On the smartphone screen, a virtual object corresponding to the recognized control marker is displayed as shown on the right side of Figure 1. Figure 2 shows markers used in the proposed system. Since this system is for Japanese learners, the character of the marker is displayed in Japanese. There are two types of markers. One is a marker to select a question, while another is a control marker that manipulates a virtual object. The virtual object acts as teaching material for mental rotation learning. The control marker is a cubic marker. When the learner rotates the control marker, a stereoscopic marker, it is possible to rotate the virtual object displayed on the screen of the smartphone. The select marker is prepared for 20 titles. The learner can switch the question by moving the select marker closer to the control marker. Figure 3 shows how the system screen looks when the question is presented. The red arrow is the reference arrow and the three straight lines form the axis. The learner thinks about which arrow among the five blue arrows...
displayed on the left corresponds to the arrow displayed on the right side of the screen. The proposed system uses two buttons in the proposed system. “Inputting answers button” is used for learning graphics using the system and transitioning to the screen to input answers. “Arrow button” is used for answering directions when entering forecast and answer. The buttons are displayed in Japanese. The learner can input the forecast of the solution to the question by selecting the arrow on the screen. After inputting the forecast, the questions are displayed in the upper left side of the screen. The learner considers his or her answer while confirming the question and the virtual object displayed on the screen. If the learner answers correctly, the system finalizes the solution to that question and the learner can move on to the next question. In case of an incorrect answer, the learner returns to input the forecast. By using their markers, the learner can learn with subjective “operation”.

3. Preliminary Experiment

In the experiment, in order to confirm the operability of this system, we instructed and evaluated learning using the proposed system. The subjects were six university students (A to F), among them three subjects A, B, and C formed Group 1 (the experiment group) and used the proposed system. The other three subjects (D, E, and F) formed Group 2 (the control group) and used a system that presents the question and performs correct/incorrect judgment. The system used by Group 2 displays a question
on the screen, and when the learner answers the question, “correct answer” or “incorrect answer” is displayed for the subjects. This feedback is also offered in Japanese.

During the verification, the learners first answer the questions on mental rotation as part of a preliminary test comprising 20 questions. The subjects then learn using the system assigned to each group. After learning using the system, learners solved the same questions that were posed in the preliminary test again, after which a post-test questionnaire was carried out. Table 1 shows the number of correct answers on each subject’s pre-test and post-test. Because questions are too easy, there was no significant difference in the number of correct answers between the two groups, as shown in Table 1.

Table 1: Number of correct answers on each subject’s pre-test and post-test.

<table>
<thead>
<tr>
<th>Group</th>
<th>Subject</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (experiment group)</td>
<td>A</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>2 (control group)</td>
<td>D</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

The post-test questionnaires show that it was easy to grasp the shape of the figure using the proposed system, and the subject could be confirmed answers by using system. It also shows that subjects could understand the questions by learning what they could not understand at first and that the system is simple to operate. On the contrary, there are opinions such as “when the marker is hidden with the finger, the recognition is interrupted and the figure is not displayed” and “It is inconvenient that one hand cannot be used.” Therefore, the subjects faced problems using the stereoscopic AR markers.

4. Summary and Future Challenges

Spatial cognitive and mental rotation abilities are considered fundamental to solve the questions on spatial figures. In the conventional system that requires learners to answer questions on spatial figures using AR, it is difficult to grasp the shape of the surface behind the marker. Therefore, in this research, we proposed a system that allows learners to answer questions on mental rotation using a smartphone and AR markers. In the proposed system, it is possible for the learner to grasp the shape of the figure from all directions while operating the cube marker.

During the preliminary experiment, we compared and verified operability of proposed system. The experimental results did not reveal any significant difference in the results from using the proposed system; however, from the post-test questionnaire, it was confirmed that the proposed system could enable learners to grasp the shape of the figure and confirm the solution better. Also, learners had opinions like, “I did not know what to solve because I did not understand knowing the intention of the question” and “There was a place where it was difficult for the arrow to tap when answering.” Therefore, in the future research, improvement of the user interface such as that for manipulation using AR marker. In addition, we consider the difficulty level of the problem and the subjects, conduct experiment which can verify the learning effect by involving subjective manipulation.

A part of this research has been made possible through Grant-in-aid for Scientific Research C (No.15K01084) from MEXT Japan.

References


Monitoring System to Help an e-Learning Institution to Manage Tutors and Student Data Retrieved from Moodle

Elza MONIER*, Fátima GATINHO, Ana Emilia OLIVEIRA, Aldrea RABELO & Rômulo MARTINS  
UNA-SUS, UFMA, BRAZIL  
*bernardes.elza@gmail.com

Abstract: E-learning has become an excellent alternative for those who want to align studies, professional qualification, and career. Several courses offered in this modality use the Virtual Learning Environment (VLE) known as Moodle to support the process of teaching and learning. However, some managerial needs are not optimized in this VLE, especially regarding the administration of students and tutors information. Because of this, the monitoring process carried out by the work teams in the educational institutions might become expensive and time-consuming. Although the data is in a single database in Moodle, it is not a practical activity retrieving the information. Therefore, in face of this reality, the Open University of Brazilian National Health System in partnership with the Federal University of Maranhão (UNA-SUS/UFMA) and other higher education institutions developed the Monitoring System, a tool that makes easier the decision-making and the didactic strategies elaboration using data from Moodle. This paper aims to present the Monitoring System, that enabled data management in a faster and online way, requiring less effort of the monitoring team to generate useful information for a satisfactory course control and progress, resulting in a significant reduction in the time spent by UNA-SUS/UFMA monitoring team to manage students and tutors data.

Keywords: Virtual learning environment, Moodle, Monitoring System.

1. Introduction

With the worldwide advancement of technology, the openness to create educational alternatives and new forms of learning has become possible, such as e-Learning modality. This educational model has arisen as an appropriate and essential alternative for a part of the population who cannot participate in face-to-face courses due to distance limitations or lack of time (Frota et al., 2015).

According to Groenwold & Knol (2013), the number of e-Learning courses is likely to increase over the coming years, as well as it is expected an equal rise in the development of Virtual Learning Environments (VLE) that are aimed at providing support during the teaching learning process in e-Learning initiatives (Maciel, 2013). Currently, Moodle is one of the most used VLE. Considering this context, this paper presents a solution for problems related to the administration of data retrieved from Moodle.

In the Northeast of Brazil, the Open University of Brazilian National Health System that works in partnership with the Federal University of Maranhão (UNA-SUS/UFMA) has experienced some difficulties in managing information retrieved from Moodle. To solve the problems, the institution has created a Monitoring System that enables professionals to process tutors and students' data collected at the Moodle in a faster and easier manner. This paper aims to present this Monitoring System as an alternative to improve the monitoring of students and tutors in e-learning courses that use Moodle as VLE, helping to prevent evasion.
2. Theoretical Background

Educational processes in distance learning when based on Information and Communication Technologies (ICT) use - educational modality known as e-Learning - are facilitated and improved through VLE, such as Moodle (Tobase, Guareschi & Frias, 2013; Hannel, Lima & Descalço, 2016).

Moodle (Modular Object-Oriented Distance Learning) is a learning platform developed for teachers and/or tutors, students and administrators as a tool for interaction, with a focus on creating personalized learning environments (Serra et al., 2016).

It is composed of several tools that enable the publication, interaction and evaluation of resources and activities in the learning process, such as Online Books, Labels, Wikis, Chats, Forums, Diaries, Questionnaires, Tasks, among others (Kraemer, 2015).

It also allows the extraction of data for electronic monitoring of the student, from the manual survey of quantitative data, such as student time in the virtual environment, what was accessed and the frequency with which an element was accessed (Schneider, Mallmann & Franco, 2015). This information, however, is diffuse in the Moodle pages, demanding more time and work from a technical team to monitor them during the courses.

According to the latest analytical report of Distance Education in Brazil, the dropout rate is approximately 25% in authorized courses (Brazilian Association for Distance Education, 2015). To reduce this indicator, it is necessary to implement monitoring strategies in a continuous way, with systematic use of prior planning, allowing the development of skills through the resolution of problematic situations in the teaching-learning process (Schneider, Mallmann & Franco, 2015; Serra et al., 2016). Monitoring is an essential task to investigate the limiting situations and the advances of teaching-learning in the distance modality.

3. Monitoring System for Moodle VLE

The Monitoring System is a complement to the academic management of UNA-SUS/UFMA Moodle, which generates detailed reports, according to the needs of the institutions of higher education. Using it makes possible the monitoring of student's notes, access time in Moodle, postings of activities, the percentage of messages exchanged between students and tutors, participations in forums, etc., qualifying the information that is relevant to the evaluation and satisfactory progress of the courses, in a faster and more accurate way.

The user accesses the system, visualizes all available courses and modules and when choosing a course or module, several types of reports are presented containing: information about students' and tutor's journals, quizzes, discussion forums, reports of student participation in the Final Paper stage, frequency of access, messages exchanged between students and tutors, general report, an area to view students' notes similar to an academic record, among others.

4. Technical View of the Monitoring System

The system was developed using PHP 5.x, HTML, CSS, JavaScript and the JQuery libraries for the interface. The database manager used is MySQL, the same manager used by Moodle. In the current version of the Monitoring System, versions 1.9 to 2.9 of Moodle are supported, and there are changes of business rules between these versions.

In addition to supporting the various versions of Moodle, it allows the integration of the data obtained from the versions in a single environment, so it is possible to consult data from several environments in a single application. This justification can best be visualized in figure 1. All access to the database query used is shared with Moodle, thus facilitating the integration of information.
5. Final Considerations

Moodle has been widely used in distance education since it enables interaction among those involved in the learning process (tutors, professors and students). It also allows the follow-up of students through educational indicators. However, it provides important data in different areas, and when it is necessary to have an overview of students this view takes twice as long, because all data extraction is done manually.

The Monitoring System integrated to Moodle, developed by UNA-SUS/UFMA in partnership with other higher education institutions, is an efficient tool for controlling the performance of students and tutors. It provides the possibility to monitor students and tutors in Moodle version 1.9 and 2.9, in a faster and online manner, requiring less effort by the monitoring team to generate useful information for satisfactory course control and progress.

With the Monitoring System the information is presented in a general overview of students and teachers, with minimal effort, compared to directly consulting in moodle. The implementation of this system resulted in a significant reduction of the time spent by the UNA-SUS/UFMA monitoring team to generate useful data. It also allowed a better and greater monitoring of the access and performance of the students and tutors in the graduate courses offered by the institution.

References


A Study on Prediction of Academic Performance based on Current Learning Records of a Language Class using Blended Learning

Byron SANCHEZA*, Xiumin ZHAOB, Takashi MITSUSHIC & Terumasa AOKID

a Graduate School of Information Science, Tohoku University, Japan
b Center for Educational Informatics, Tohoku University, Japan
c Institute for Excellence in Higher Education, Tohoku University, Japan
d Graduate School of Information Science, Tohoku University, Japan

*byronism@riec.tohoku.ac.jp

Abstract: In this paper, we describe a classification method that does not rely on historic data to predict changes in student academic performance, and therefore predict if a student will fail a class or not. By classifying students into groups given their grades, and extracting the common features in between them, it is possible to use those common features to predict if other students that share common characteristics will fall into the same classification groups. As well, those same common features can be used to help students improve their academic performance.

Keywords: K-Means clustering, feature selection, student performance prediction, unsupervised learning, learning analytics

1. Introduction

Learning assisting systems and educational models are ever growing areas in the world of education, and as so, topics of great interest. The following is part of a bigger 3-phased learning process designed to create an educational learning system that promotes continuous and sustainable learning. This educational environment includes a micro-learning environment and a smartphone application. Both designed to help struggling students in class by providing teacher to student feedback, and after-class learning exercises created to help with the processing of new lessons and information. One of the biggest challenges with this approach is being able to identify students in need of guidance, and even more so, predicting which ones are prone to having problems in a near future.

Nowadays, a large amount of research regarding learning analytics and the prediction of student performance using data mining is readily available. Ueno (2003) and Pardo et al. (2016) created predictive models using large amounts of historic data in order to predict student performance. Ade and Deshmukh (2016) demonstrated that by combining the expected outcomes of two well know algorithms using different voting strategies, a 3% increment to the predictive accuracy level of the next highest performing algorithm could be obtained. Bote-Lorenzo and Gómez-Sánchez (2017) and Hlosta et al. (2017) both recently made predictions using the information recollected up to the current lesson as model data to predict engagement and dropout risk, and therefore are key to this research.

In all these studies, student performance was successfully predicted at different accuracy rates using high volumes of data. Unfortunately, that is not the case in many educational scenarios, for which we apply a voting scheme combined with current data to perform predictions and make up for the low volume of data available.

2. Proposal of Prediction Method

The following method consists of a four-step linear process that utilizes K-Means clustering to create potential classification of students based on their academic performance, and then tries to predict
changes in those groups throughout the lessons using a decision tree algorithm (the summary of grades represents academic performance in the context of this paper).

First in the process, the data must be prepared to perform the analysis and clustering of such. Normalization was performed using the Z-Score for scaling, and new features were added to tell us how their values have changed throughout the lessons. The following 2 formulas were used for each feature:

\[
PI = \frac{v_k - v_{k-1}}{\min(v_k, v_{k-1})} \times 100, k > 1 \quad (1)
\]

\[
API = \sum_{k=2}^{n} \left( \frac{v_k - v_{k-1}}{\min(v_k, v_{k-1})} \times 100 \right) \quad (2)
\]

In formulas (1) and (2), \(v_k\) and \(v_{k-1}\) are the current and last lesson’s grade averages.

The second step consists of classifying students by their academic performance in class using only features that represent grade performance. By performing clusters of students based on their academic performance, it is possible to separate them by how similar their grades behaved throughout the lessons taken. We used a K-Means clustering algorithm, with four centroids and a K-Means++ initialization step, using the Euclidean Distance as a distance measurement.

In the third step, we use different combinations of the previously extracted features in order to create models that best identify the current data, and then use those models to predict the next classification of students. Models were created by combining a set of features, an amount of past lessons of information, and whether they include the previous cluster results or not. We rank the models by their accuracy score and choose the models that have the highest values as prediction models for the next lesson.

The final step of the process uses the before created prediction models, and applies the data from the last lesson as input in order to predict the classification groups of the next lesson. Given that there are many models possibly with the same accuracy value, we choose all models that have a value above the 3rd highest accuracy score, and created a voting scheme were the most voted group by all models is the decided group. We proceed to do this entire process for each lesson predicted.

3. Experiment and Results

The data used belongs to a Chinese - Japanese language course. The class contains 32 students, and 12 lessons. Grades are scored over 100 points, and students may attempt a same exercise an unlimited amount of times. The features obtained from the data are: "Score", "Finished attempts", "Abandoned attempts", "Time between exercise release and first attempt", "Avg. time to complete an attempt", "Time between first and last attempt".

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Top scoring model</th>
<th>Voting scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 6</td>
<td>0.8125</td>
<td>0.90625</td>
</tr>
<tr>
<td>Lesson 7</td>
<td>0.78125</td>
<td>0.96875</td>
</tr>
<tr>
<td>Lesson 8</td>
<td>0.875</td>
<td>0.90625</td>
</tr>
<tr>
<td>Lesson 9</td>
<td>0.8125</td>
<td>0.90625</td>
</tr>
<tr>
<td>Lesson 10</td>
<td>0.6875</td>
<td>0.9375</td>
</tr>
<tr>
<td>Lesson 11</td>
<td>0.875</td>
<td>0.875</td>
</tr>
<tr>
<td>Lesson 12</td>
<td>0.8125</td>
<td>0.9375</td>
</tr>
<tr>
<td>Average total</td>
<td><strong>0.8203</strong></td>
<td><strong>0.9178</strong></td>
</tr>
</tbody>
</table>

When using the newly added features in combination with the existing data for student grades, the clustering algorithm correctly identified four groups of students. Figure 1 shows the academic performance of each student as a line, and to what classification group the student belongs to as color (data was used up to lesson 5 as an example). Following steps 3 and 4, we obtained a list of the highest scoring attribute/parameter combination and used those to create prediction, and then used the models to predict on what group to classify each student on each lesson. Table 1 shows the result comparison.
between using the voting scheme and just the top scoring model (the first lessons haven been omitted because the values do not change significantly).

The results improve when using a voting scheme in most cases. We also found that using the top 3 highest accuracy values, gives us the best results. Models used after the 3rd do not increment the accuracy significantly. These results were omitted.

![Figure 1. Student academic performance groups](image)

4. Conclusions

The results found in this paper indicate that by applying this method to a set of features, we are able to predict with a high accuracy the changes to student performance in a class with limited amount of data. By identifying/predicting which groups have declining or improving grades, and helping students in those groups, overall academic performance can be monitored, and most of all improved with proper management. As future work, we shall try different prediction algorithms and different data from another year of the course, with the purpose of improving accuracy and validating the method.

Acknowledgments

This work was supported by the JSPS KAKENHI grant, numbers JP15K02709 and JP15K01012. We would also like to thank the members of both Mitsuishi Lab and Aoki Lab at Tohoku University for their insight and input into the topic since the beginning.

References


The Development of a Simulation to Support Authentic Observation in Precipitation Reaction

Chang Youn LEE & Hun-Gi HONG*
Department of Chemistry Education, Seoul National University, South Korea
*hghong@snu.ac.kr

Abstract: In this study, we developed an interactive simulation which can provide authentic observation experience on precipitation reaction by embedding video clips of the actual experiment result. The purpose of this development was to assist students in more authentic observation and modeling on the precipitation reaction. The simulation was built with Flash and ActionScript to runs on wide-ranging devices regardless of operating system. The development process of a simulation through design seminars to reflects the diverse perspectives of teachers and science education experts. Then, we confirmed the validity of the development through the expert evaluation of two experts of science education and four teachers who did not participate in the design seminars. Attempting to introduce new characteristics into science simulation can be attributed to extending the field of simulation design on science education. This study is thought to be useful to come up with a design form suitable for science digital textbooks.

Keywords: Simulation design, Science simulation, Flash, ActionScript, Precipitation reaction.

1. Introduction

Precipitation reaction, as educational experiment in middle school, allows students to understand existing ions in transparent aqueous solution. As the reaction proceeds, a characteristic color of precipitate is observed gradually with naked eye, which helps students with constructing a scientific model of invisible ions. K-12 students who are not accustomed to abstract thinking need an opportunity to observe a continuous process of actual phenomenon in order to construct the ion model effectively (Harrison & Treagust, 2000). However, if the school does not provide laboratory activities due to time, space, or economic constraints, most of students have to rely entirely on textbooks to experience experimental activities indirectly (Park, 2011). Traditional textbooks carry static graphical information, so that they rarely allow students to observe dynamic process of the precipitation reaction (Smetana & Bell, 2012). Furthermore, school culture, which tends to emphasize explicit knowledge for testing, has often emphasized linguistic information, such as color, rather than the phenomenon of precipitation itself (Clark & Mayer, 2016; Wong, Milrad, & Specht, 2015). Emphasis on this kind of well-made scientific knowledge can impress students as a monotonous knowledge to learn, while at the same time hampering higher-order thinking (Jonassen, 2010).

As a candidate to address these issues, simulation can be used as a supplementary material, even in the didactic instruction. A number of science simulations that have been developed so far show dynamic changes of ideal scientific phenomena based mainly on the animation, but the development of a science simulation embedded with video clips has yet to be attempted (Rutten et al., 2012).

In this study, we focused on a design of the science simulation that intend to assist the students in observation as well as modeling that has already attracted attention in precedent literature(Rutten et al., 2012; Smetana & Bell, 2012). Concretely, we developed an interactive simulation that allows student to experience authentic observation on precipitation reaction by embedding video clips of the experiment results. It is built with Flash and ActionScript so as to make it run on various kind of devices regardless of operating system.
2. Methodology

The development process of a simulation was carried out through design seminars, which follows the characteristics of the participatory design introduced by Könings (2014). Design seminars was attended by three chemistry teachers and one professor of chemistry education, in which they provide design comments based on their educational context. Könings (2014) consider that the diverse perspectives of developers, field teachers, and students can reflect by letting them participate in the design process as a designer. But, since students’ perspectives are interested-oriented, not professional, we did not consider their participation (Cober et al., 2015). In this study, the researcher take over the role of developer, and held seminars to discuss a design of the simulation prototype with the participants. Then, we evaluated the simulation based on the method of the expert evaluation with two experts of science education and four teachers who did not participate in the design seminars.

In the early stages, the design of the simulation was suggested by the researcher’s heuristic knowledge and complemented by participants’ comments in the design seminars as well as well-known design principles including the multimedia design principles of Mayer (2005). The simulation developed in this study has two characteristics. First, a video clip of the experiment results taken by the researcher was inserted to express a dynamic process of the precipitation reaction on the reaction plate. At this time, we used a mask function to naturally match square-shaped movie clips with reaction plate graphic objects. The mask function refer to a function that transforms the boundaries of an original object by displaying only the overlapping parts between the original raw object and the freely drawable shape (Figure 1). This minimizes the unnaturalness in layout of screens between different representations (video, animation, graphic).

![Figure 1. Transforming the movie shape through mask function](image)

Second, an interactive activity, which contains a scientific model of the invisible ion on precipitation reaction was provided. For example, when mixing solutions in two beakers in a simulation, a dynamic model of the precipitation reaction process is shown in a new beaker (Figure 2 (b)).

3. Results

A development result of the simulation on precipitation reaction is presented in Figure 2. First, a scene example to support authentic observations of precipitation reaction is shown in (a). In the screen, round buttons including arrow are provided for students to drop the reagent interactively asynchronously. In this way, a total of four sets of experiments are provided in this simulation. The nameplate of each aqueous solution is colored as cyan and red to help students differentiate cations and anions from the chemical formula. Second, a scene example to support modeling is shown in (b). It allows learners to build models based on their precedent observation experience in the simulation. If one of the two beakers is pulled and overlapped, the animation of a dynamic model on precipitation reaction is played. In the animation, precipitate is created when two ions involved in the reaction meet each other. The learner can select another precipitate and perform modeling exercises on it. This simulation provides dynamic models of nine precipitate including CaSO₄, BaSO₄, CaCO₃, BaCO₃, AgCl, PbI₂, CdS, PbS, and CuS. Meanwhile, analysis of the expert ratings is now in progress.
4. Conclusions

This study presents a case of a science simulation embedded video clips showing practical results in the precipitation reaction. In the development, we focused on a simulation design to provide students with more authentic observation experience than information provided just by the animation. In addition, the simulation incorporates modeling activity based on inherent interactive characteristic of Flash to support deep understanding of presence of invisible ions.

Attempting to introduce new characteristics into science simulation can contribute to extending the field of simulation design on science education. With the recent introduction of classroom technology, simulations are increasingly being included in digital textbooks. This study is thought to be useful to come up with a design form suitable for science digital textbooks.

Acknowledgements

This study was developed from the extension of a Lee’s Master thesis (2015) and reconstructed to introduce design characteristics of the science simulation.

References


Kanji Learning Support with Feedback based on Haptic and Pseudo-Haptic

Takanori Kono*, Yukihiro MATSUBARAa & Masaru OKAMOTO*
*Graduate School of Information Sciences, Hiroshima City University, Japan
*lkono@lake.info.hiroshima-cu.ac.jp

Abstract: Learning systems with haptic or pseudo-haptic feedback has developed. However, a study have not been conducted about effect of haptic and Pseudo-haptic feedback in learning system. Therefore, in order to investigating effect of haptic and pseudo-haptic feedback in learning system, we developed Learning system with feedback based on haptic and Pseudo-haptic.

Keywords: Force feedback, Pseudo-Haptic feedback, learning support system for Kanji.

1. Introduction

Recently, various learning support systems using haptic devices were developed. Haptic devices can give users force feedback from manipulators, which is one of their uses. In learning support systems, force feedback is used for various purposes. Hidani et al. (2015) developed the virtual experiment environment for pulley learning using a portable haptic device (SPIDAR-tablet). SPIDAR-tablet is a haptic device for tablet type PC (Tamura, Murayama, Hirata, Sato, and Harada, 2011). Isshiki et al. (2009) used haptic interaction for English-accent learning. As another force-experience approach, Pseudo-Haptic is proposed (A. Lecuyer, 2009). Shiota et al. used Pseudo-Haptic for feedback to create a concept map and evaluated the effects of it (Shiota and Kashihara, 2013). Yamasaki et al. showed the training effects of using Pseudo-Haptic in motion training (Yamasaki, Hanai, Arai and Seo, 2015).

Comparisons of effects of haptic feedback and the effects of Pseudo-Haptic feedback in learning support have not been conducted. In this paper, we developed a PC tablet-based learning support system using real force feedback and Pseudo-Haptic feedback. We also compare effects of haptic approach and Pseudo-Haptic feedback approach, as well as sound feedback in learning support systems. We then conducted an evaluation experiment with this novel system. In this study, a prototype system was developed for learning stroke order of Kanji (a kind of character set used in Japan).

2. Learning system for Kanji (Characters used in Japan)

2.1 Outline of the System

This system supports learning of stroke order of Kanji (characters used in Japan). The system uses three methods (Force, Pseudo-Haptic, and Sound) to give learners feedback about stroke order. In order to compare each feedback method, the system points out mistakes by a single feedback method. Learners enter stroke order by writing Kanji displayed on the tablet PC screen. The system instructs learners by selected feedback method if and when a learner makes a mistake. The system judges stroke order by start-point coordinates, endpoint coordinates, and by node coordinates.

Figure 1 shows the overview of the system. The system is constructed with a tablet PC and a SPIDAR-tablet. Force feedback of the system is done with the SPIDAR-tablet which is constructed with a black frame, motor, string, and ring as seen in Figure 1. Learners fit the ring (see Figure 1 center), which is connected with strings. The system performs force feedback by controlled tension of strings by the motor. SPIDAR-tablets are used when force feedback is selected as the feedback method.
2.2 Method of Pointing out Errors

In this study, incorrect stroke orders are classified as “Order Error”, “Short Stroke”, “Over Stroke”, and “Reverse Stroke”. The system points out errors to the learner to help the learner distinguish the type of error; however, the system points out “Order Error” and “Reverse Stroke” by the same method. Therefore, the learner distinguishes 3 types of errors. Pointing out errors by sound is conducted by the same buzzer. However, the learner can distinguish types of errors by timing.

Here we describe each type of error. Figure 2 shows an example of each type of error. In Figure 2, the black painted stroke corresponds to the written stroke; red arrows are incorrect strokes; and, green arrows are correct strokes as well. “Order Error” is an error that learners wrote an incorrect stroke. “Short Stroke” is an error that the learner took a finger off at a point, which was before the correct end point. “Over Stroke” is an error that the learner wrote another stroke beyond the correct stroke. “Reverse Stroke” is an error that the learner wrote from the end point to the start point.

Here we describe concrete method of pointing out. If the learner commits an “Order Error” in the haptic mode, the system provides force feedback in the direction of the point that the learner began writing. In the Pseudo-Haptic mode, when the learner touches the screen, the system begins to provide Pseudo-Haptic. Next, we describe the method of pointing out “Short Stroke”. In the haptic mode, when the learner takes the finger off, the system provides force feedback in the direction of the point that the learner began writing. In the Pseudo-Haptic mode, when the learner takes the finger off, the system moves the pointer to point that the learner began writing. Next we describe the method of pointing out “Over Stroke”. In the haptic mode, when the learner writes away from the correct ending point, the system provides force feedback in the direction of the correct ending point. In the Pseudo-Haptic mode, when the learner writes away from the correct ending point, the system provides Pseudo-Haptic.

In the sound mode, the system points out each error by the same buzzer, however, the learner can distinguish the type of error by timing that is the system begins buzzer.

3. Experiment

3.1 Purpose and Method

This experiment was carried out with the system to investigate difference of learning effects that arise from each feedback method and the peculiarity of each feedback method in the learning system. Nine Japanese subjects (4 college students and 5 graduate students) participated in the experiment. The subjects were first tested on 24 questions about the stroke order of Kanji that is easy to mistake as pretest (paper test). After that, the subjects learned 8 Kanjis by each of the feedback methods. After learning, a post-test was conducted in the same way as the pretest was conducted, and we investigated the learning effect. Finally, a questionnaire was conducted.
3.2 Results and Consideration

Figure 3 shows the correct answer rate of pretest and post-test. Figure 3 indicates the amount of change from pretest results to post-test results, where each feedback method was nearly unchanged. Figure 4 shows the average touch count to users writing a correct stroke after a learner commits “Over Stroke” with sound feedback equates to more counts than with force feedback and Pseudo-Haptic feedback. We confirmed that touch count to users writing a correct stroke after a learner commits “Over Stroke” with sound feedback equates to more counts than with force feedback and Pseudo-Haptic feedback. We considered that learners find it difficult to distinguish the type of error by sound feedback, and that learners made the same mistake after the system pointed out error. The questionnaire also helped confirm a tendency that learners cannot distinguish the type of error by sound feedback. We confirmed personal difference in degree of felt of Pseudo-Haptic feedback is larger than force feedback, and users who hardly feel Pseudo-Haptic feedback find it difficult to distinguish the type of error. Some subjects who hardly feel Pseudo-Haptic feedback grasped it as a visual effect and thus could distinguish the type of error.

![Figure 3](image1.png)  
**Figure 3.** Correct answer rates of pretest and post-test.

![Figure 4](image2.png)  
**Figure 4.** Average touch counts until users write a correct stroke after a learner commits “Over Stroke”.

4. Conclusion

In this study, we developed a learning system for Kanji that can provide force feedback, Pseudo-Haptic feedback, and sound feedback. The experiment confirmed that learners sometimes find it difficult to distinguish the type of error by sound feedback; additionally, we confirmed personal difference in degree of felt of Pseudo-Haptic feedback is larger than force feedback. We need to investigate a learning subject that can confirm the differences of learning results across each feedback method. Part of this research was made possible through the Grant-in-Aid for Scientific Research C (No.15K01084) from the MEXT Japan.

References


Discourse Analysis of Teachers' Commentary on Students

Rui Shi*, Longwei Zheng & Xiaoqing Gu
Faculty of Education, East China Normal University, China
*yyshirui@163.com

Abstract: In this paper, we describe the application of sentiment analysis and topic model on abundant teachers’ comments. The goal of the analysis is to explore the teachers’ sentiment and concerns when making comments. Preliminary findings indicate that students’ class performance, completion of homework and subject literacy are still the main evaluation indicators which cannot keep pace with assessment of 21st century skill. Meanwhile, we proposed teachers to make more pluralistic comments for encouraging students’ development.

Keywords: sentiment analysis, topic model, 21st century skill, teachers’ comments

1. Introduction

21st century skills is an overarching expression for the knowledge, skills, and dispositions seen as prerequisites for success in the future global workplace (Germaine et al., 2016). It has been in heated debate about how to reform the curriculum, teaching method, or assessment system (Voogt et al., 2013). Many results have shown that student evaluation raises their achievements (Black & Wiliam, 2010). However, few researches have delve into in-class teaching activity to see the condition of evaluation.

To have a glance at the current situation of student evaluation, massive teachers’ comments have been investigated by natural language processing method. We answer the following research questions:

1) How to infer teachers’ emotions by their comments?
2) How to reduce teachers’ focus when evaluating?
3) How to interpret these indicators with 21st century skills?

The 21st century skills in this study are referred to use the framework put forward by the National Education Association (2012) as the 4Cs (Germaine et al., 2016): 1) Critical thinking and problem solving skills; 2) Communication skills; 3) Collaboration skills; 4) Creativity and innovation skills.

2. Data and Methodology

Teachers’ comments come from a student evaluation system across 8 primary schools in Shanghai and 2224 students. 18 subjects are involved in this dataset, such as Chinese, English, Math, Science and etc. 991979 comments are recorded in the database. Comment here refers to short text in Chinese.

For sentiment analysis, we firstly employed the python library ‘jieba’ to segment the words. Two corpus are built by hand to distinguish the emotion. Meanwhile, some dictionaries of adverbs of degree are included (Dong, Z., & Dong, Q., 2000). Then, we set different weights to these adverbs (Table 1), thus we can calculate the score according to the degree.

Table 1: The weight allocation.

<table>
<thead>
<tr>
<th>Types</th>
<th>Examples</th>
<th>Weight</th>
<th>Types</th>
<th>Examples</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>'most' type</td>
<td>Fully, deeply, and etc.</td>
<td>2</td>
<td>'ish' type</td>
<td>A bit, fairly, and etc.</td>
<td>0.5</td>
</tr>
<tr>
<td>'very' type</td>
<td>Much, badly, and etc.</td>
<td>1.5</td>
<td>'insufficiently' type</td>
<td>Light, slight and etc.</td>
<td>0.25</td>
</tr>
<tr>
<td>'more' type</td>
<td>More, such and etc.</td>
<td>1.25</td>
<td>'inverse' type</td>
<td>Not, none, and etc.</td>
<td>-1</td>
</tr>
</tbody>
</table>
Probabilistic topic model is an algorithm whose aim is to discover the hidden thematic structure in large archives (Blei, 2012). Here, a python library is adopted to process these comments. It mainly includes two parts, convert the segmented words to vectors and Latent Dirichlet allocation (LDA) (Blei et al., 2003). Three topics were generated after displaying of topic model for each text (positive and negative). In addition, this method is also implemented on different subjects’ evaluation discourse to explore the diversity among subjects.

3. Results

As Figure 1 shows, frequently used words in whole evaluation discourse are ‘conscientious’, ‘listen’, ‘take class’ and others, from which can be concluded that most teachers focus on students’ class performance (speech, presentation, participation). Meanwhile, the term ‘handwriting’ is frequently used which indicates writing legibly can leave a good impression on teachers. Actually, a few words can be related to the framework of 21st century skill, such as ‘presentation’, and ‘participation’.

![Figure 1](left) High frequency words in general (left). Three topics and the high frequency words in each topic (right)

In the positive text (red bubbles in Figure 2, left), high frequency words are ‘listen’, ‘read’, ‘performance’, and etc.; In the negative text (green bubbles), high frequency words are ‘speech’, ‘listen’, ‘homework’ and etc. Thus, there is no significant difference with the general text in wording. However, the difference between the positive and negative text can be concluded as the quantity and richness. In the dataset, 958,442 comments are tagged with positive label, the rest 33,537 comments are negative, which illustrates teachers are inclined to make more encouraging evaluation. As for the richness of comments, teachers make subtler expression, such as ‘participation’, ‘presentation’, ‘performance’ in positive text.

![Figure 2](left) High frequency words in positive comments and negative comments (left). High frequency words in different subjects (right).

Discourse varies from the subjects which indicates different focuses when teachers making comments. In Figure 2 (right), we show the frequently used terms in three topics with different subjects. For Chinese and English class, teachers focus on fostering students’ listening, speaking, reading and writing ability. In math classes, some new words raise up, like ‘explore’, ‘cooperation’, and others. The
word ‘cooperation’ can refer to the collaboration skills in 4Cs, ‘explore’ can be part of critical thinking and problem solving skills; In other subjects (Science, PE, Nature, etc.), words about class performance and attitude are frequently used, which may imply that teachers in these classes pay more attention to classroom rules.

Figure 3. Three topics in Math teachers’ comments

As shown in Figure 3, three topics are generated by topic model with math teacher’s positive comments. In the first topic, except sentiment words, ‘explore’, ‘calculation’ frequently arise which we can name this topic as math ability; for the second topic, the frequency of ‘practice’ are prominent, followed by other words, like ‘communication’, ‘cooperation’, thus this topic can be described as communication and cooperation ability; In topic 3, we found ‘listen’, ‘conscientious’, ‘active’ are the most frequently used words, so that we can label this topic as learning attitude. Therefore, teacher evaluation discourse in math subject mainly involves three parts: math ability, attitude, communication and cooperation ability. With this method, all subjects can be summarized into three topics which will facilitate the extraction of main information from abundant content.

4. Conclusions

In this study, we investigated teachers’ comments to be aware of current situation of student evaluation. Results are discussed from three aspects. In general, class performance, completion of homework and subject literacy are still the main evaluation indicators in most class. Meanwhile, teachers are found to make more encouraging (positive) comments to student. In comparison with other subjects, Math is a more comprehensive discipline which manifests the collaboration skills and critical thinking and problem solving skills in 4Cs. However, a large quantity of comments did not follow the trend of 21st century skills. In this context, some advanced criterion should be set to make pluralistic evaluation. Teachers can build new evaluation system based on the 21st century skills framework.

References

Discovering Teachers’ In-Class ICT Usage With Frequent Closed Sequence Mining

Bingcong WU, Longwei ZHENG & Xiaoqing GU
East China Normal University, China
wubingcong.1015@gmail.com

Abstract: The adoption of E-textbook has offered a new opportunity to gain insight into teachers’ usage of ICT in class. Usage of technology can vary greatly from one group of teachers from another and they have different using patterns. In this study, we discuss three different using patterns which we discovered with Frequent closed sequences mining. We find that different using patterns show teachers’ using habits, and can probably affect the teaching efficiency due to learners’ different learning experience in class.

Keywords: Visual analytics; Learning analytics; Frequent sequences mining; ICT

1. Introduction

Information and Communication Technologies are becoming increasingly pervasive in education (Martinovic & Zhang, 2012). It’s making a difference in the ways teacher plan lesson and organize activities. However, teachers need support to make effective use of information technology in their teaching. We did a research to discover teachers’ different usage patterns and created Sunburst graphs to provide a better understanding of the frequent sequences of visual representation and patterns. The purpose of this study is to analyze different groups of teachers with different operating behaviors based on behavior data with the intent to create a visualization of usage patterns that will help describe teachers’ in-class ICT usage.

2. Method

2.1 Data source

We investigated issues within the context of data from an E-textbook. We obtained data on teacher teaching action records, for 346 teachers enrolled in this E-textbook, observed about a month since February 16th 2017. Until March 14th 2017, the teachers have performed a total of 10424 actions which can be classified into five different actions: Lead-in class(L), Extend knowledge(E), Start New-learning(N), Draw symbols(D) and Practice(P). Typically, teachers might create a new lesson session on the system with action ‘Start New-learning’, and end the session with action ‘Practice’, these are the two basic functions of the E-textbook. Teachers can perform action ‘Lead-in class’ to help learners start learning better and action ‘Extend knowledge’ allows teachers to introduce more knowledge which is related to the lesson. They can also perform the action ‘Draw symbols’ to remind learners what is important or leave some comments to give feedback in the lesson.

2.2 Frequent Closed Sequence Mining

Sequential pattern mining has become an essential data mining task with broad applications. In this study, we used the Frequent Closed Sequence Mining, one of the methods of Sequential pattern mining which is based on the BIDE (BI-Directional Extension) algorithm to discover the teachers’ usage patterns. BIDE can be easily adapted to mine frequent closed sequences of subsets of items and it leads to a more compact yet complete result set (Wang, Han, & Li, 2007). We grouped teachers’ actions by lessons and days to generate a dataset of sequences of items. And then the data was processed with BIDE algorithm to generate a set of frequent sequences. We analyzed the features like length, occurring frequencies and complexity of these frequent sequences to discover teachers’ different usage patterns.
3. Result

A support of a sequential pattern is the number of data set sequences in which it occurred. For this research, we fixed the minimum support threshold in 50 times to consider a pattern as frequent.

Table 1: Different using patterns derived from frequent closed sequence mining

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Name</th>
<th>Typology Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Limited use</td>
<td>A few actions (less than 8 times)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Using patterns are simple (only 3 different frequent sequences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Only short frequent sequences (3 sequences in total)</td>
</tr>
<tr>
<td>B</td>
<td>Moderate use</td>
<td>Plenty of actions (8 to 38 times)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Using patterns are common (many different frequent sequences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A few mid-length frequent sequences (84 sequences in total)</td>
</tr>
<tr>
<td>C</td>
<td>Intensive use</td>
<td>Lots of actions (more than 38 times)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Using patterns are complex (lots of different frequent sequences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lots of long frequent sequences (549 sequences in total)</td>
</tr>
</tbody>
</table>

Figure 1. Using patterns

We divided teachers’ usage patterns into three groups according to the times they perform action on the E-textbook. For the top 27%, we call their usage patterns ‘Intensive use’; For those between 27% to 72%, we call their patterns ‘Moderate use’; And the other teachers’ usage patterns are ‘Limited use’, they operated less than 8 times.

Pattern A is ‘Limited use’, we discovered that they only have three different frequent sequences, which are \{N\}, \{P\}, \{N, P\}. As shown in figure 1, this using pattern is very simple. Typically, when start a new lesson, teachers would perform the action ‘Start New-learning’ and they might end the
lesson with action ‘Practice’. They probably didn’t use the E-textbook during other periods of the class, so we assume that they were not using the E-textbook frequently and effectively to bring better learning experience to the learners.

Pattern B is ‘Moderate use’, teachers with this using pattern operated all five actions on the E-textbook as they tried to explore all the different functions of the E-textbook, so it has many frequent sequences with different combinations in different lengths, which makes pattern B more complex than pattern A. As shown in the Figure 1, pattern B has some mid-length frequent sequences like \{N, E, N, P\}, the longest frequent sequence of pattern B is \{N, P, E, N, P, E\}, a 6-length sequence with a support of 53. With these frequent sequences, we assume that teachers with pattern B were making effort to discover a better way to use the E-textbook in their teaching and create better learning experience for the learners. For example, the frequent sequence \{N, P, E, N, P, E\} shows that teachers with pattern B tried to start more than one action ‘Start New-learning’ in a class, they started new learning for twice in a class, and after each learning they would assign some practices for the learners to consolidate knowledge. Also, after each practice, teachers would perform action ‘Extend knowledge’ to provide learners more knowledge which is related to the new learning and the practices to help learners understand the lesson better after they finish the practice.

Pattern C is the ‘Intensive use’, a very complex and diverse using pattern. Some of the frequent sequences can be very long like \{N, L, N, D, E, N, P, E\}, an 8-length sequence. After started a new learning at the beginning of the class, teachers were trying to use attractive methods or materials to lead-in the learning in a much more interesting way. With the action ‘Lead-in class’, a class could be more interesting, learners probably felt excited about the learning and had a better learning experience, action L is quite important at the beginning of a vivid class. After finished the first learning period of this class, teachers started another new learning period with another action ‘Start New-learning’, then they performed action ‘Draw symbols’ to draw symbols or leave some comments to remind the learners to prepare for learning more extensive knowledge which was showed to learners with action ‘Extend knowledge’. Teachers even started the third new learning in this class, after the first two learning, teachers skillfully performed action ‘Practice’ to consolidate the knowledge the learners had in the class. Finally, teachers performed action ‘Extend knowledge’ to offer learners chances to discover more by themselves and make decisions about what was the most they want to learn about this class. With a total of 549 sequences and half of them are 4-length sequences or even longer sequences, and many of these frequent sequences had supports larger than 100, we can assume that teachers with pattern C tried to perform different actions to seek a better way to use the E-textbook in class. In conclusion, teachers with pattern C tried so many complex combinations to improve their teaching and E-textbook using skills, they are very intensive to use the E-textbook in different ways to help them to teach more effectively and bring better learning experience to learners of their class.

4. Discussion

This study introduced a real-world situation of some teachers’ E-textbook usage patterns. We discovered three different usage patterns and assumed that different patterns might represent different efficiency of using the E-textbook in class. Typically, we assume that teachers with pattern which is more complex can probably bring more resources to the learners in the class to make their lesson more attractive and effective. And with some more rational combinations of actions in their teaching with E-textbook might make the whole process of their teaching more reasonable with higher efficiency. Our future work will be oriented to discover how these different kinds of ICT patterns would affect teachers’ outcome. We want to be able to provide more advises for teachers to improve their teaching efficiency with better ICT patterns.

References


Wang, J., Han, J., & Li, C. (2007). Frequent closed sequence mining without candidate maintenance. IEEE Transactions on Knowledge and Data Engineering, 19(8)
Support for the Cycle of Task Extraction, Goal Setting and Assessment in Research Activities

Shigeki OHIRAb*, Naoya MORITAb & Katashi NAGAOa

aInformation Technology Center, Nagoya University, Japan
bGraduate School of Information Science, Nagoya University, Japan
*ohira@nagoya-u.jp

Abstract: We previously developed a research activity concierge (RAC) system, which is a platform to encompass general research activities, and applied gamification to this system to keep user motivation high. However, even with the RAC, non-research-savvy students have difficulty executing tasks. In this research, we introduced a mechanism to support task execution in research activities by implementing automatic task extraction into the RAC.

Keywords: Research activity, task execution, automatic extraction, gamification, seminar

1. Introduction

Research consists of various activities. However, it can be seen that the everyday activities are very mundane. Scientists must often carry on without immediate visible results. On the other hand, gamification—the use of game design elements in non-game contexts (Deterding, Dixon, Khaled and Nack, 2011)—has attracted enormous interest across a range of different areas, including education (Cronk, 2012; Kotini and Tzelepi, 2015). We have developed a comprehensive gamification framework for research activities (Ohira, Sugiura and Nagao, 2015; 2016). Specifically, it provides a research activity concierge (RAC) system, which is a platform for recording and organizing everyday challenges and tasks arising in discussions and for visualizing the results when they are applied to real actions.

However, it is difficult for non-research-savvy students to accurately handle day-to-day challenges and completely execute day-to-day tasks. In this research, we focused on the discussions in seminars and introduced a mechanism to support task execution in students’ research activities by implementing automatic extraction of task statements into RAC. Moreover, we conducted experiments to quantitatively evaluate the effect on research activities of the proposed RAC and a qualitative assessment by a questionnaire. The paper will report the results of supporting task execution.

2. Research Activity

The IDC (interest driven creator) theory (Wong, Chan, Chen, King and Wong, 2015) has recently been attracting much attention, and learning and teaching methods have been changing from “examination-driven” to “interest-driven,” that is, students study what interests them. Also, to fully engage in research activities, researches must be interested in them. However, students often do not know what kind of activity to carry out because they have not been shown a global image of research. Therefore, it is thought that research activities can be more smoothly executed by preparing guidelines of research activities for such students. We have classified all research activities into 11 main activities and 100 sub-activities. We call this the research activity map and express it in the mind map format.

Challenges and tasks also need to be clear for research activities to be smoothly executed. In particular, students just starting their research have difficulty setting appropriate challenges and tasks on their own, and discussion within the research group is extremely important in resolving this. We have been developing and operating a system to record seminar content including videos, slides, and a summarized transcript. After a seminar, students review and organize what was discussed. Reviewing seminar content takes effort, so a mechanism is needed for focusing on salutary opinions and advice for subsequent activities and actively promoting the recording and organizing of challenges and tasks.
3. Research Activity Concierge

The research activity concierge (RAC) is a comprehensive support system for general research activities that introduces a gamification framework for organizing challenges and tasks on the basis of seminar content and visualizing research activity on the basis of the performance of the tasks. The research activity concierge system consists of three basic tools: the research activity organizer (RAO), the research activity visualizer (RAV), and the research activity watch-dog (RAW). RAO organizes challenges and tasks on the basis of seminar content, and RAV visualizes research activity on the basis of the performance and tasks. RAW is a tool that constantly monitors the information input and output from RAO and RAV, describes the current status of activity to the user, and recommends actions.

4. Task Execution Cycle

The task execution cycle consists of three phases: “task extraction” using seminar contents, “goal setting” based on tasks, and “assessment” of task achievements.

4.1 Task Extraction using Seminar Contents and Goal Setting based on Challenges and Tasks

Although looking again at seminar content including useful discussions for research activities is important, it also requires time and effort. Thus, task statements have been automatically extracted by using a machine learning model that has metadata in a seminar content and linguistic information of the utterance as features (Nagao, Inoue, Morita and Matsubara, 2015).

A user judges whether or not the statement extracted automatically is really a task statement and clicks the appropriate button on a statement list in the statement view of RAO. The task/non-task statement information is fed back as a teacher signal and is used for active learning (Settles, 2010) that updates a task statement extraction model (Figure 1). After judging a task statement, a user quotes the fixed task statement to a note view and creates a memo. Four types of progress tags can be attached to a created memo: not-started, in-progress, completed, or pending. Moreover, target actions can also be selected from elements included in a research activity map and attached to memos in the form of tags. Thus, created memos can record the research activity to which they are related.

A student with little experience of research activities often feels uncertain about how long to spend executing one of several tasks. Therefore, the time spent on each task should be managed. In this research, we introduced a scheduler for task execution as a function of RAV. In the task scheduler, graphs and a calendar are arranged, and users can see the approximate ratio and achievement status of tasks. Moreover, users can schedule task executions on a timetable. Users can plan a reasonable task execution schedule by checking their free time. The free time is calculated by subtracting the scheduled time of all the tasks from the maximum activity time set beforehand. Since RAV sets a rule-of-thumb achievement time of a target action on the basis of results of a questionnaire to students, the simple automatic scheduling function is also implemented in the task scheduler.

4.2 Recording and Evaluation of Task Achievements

After task execution, users record details, such as the contents of execution, on the memo that quotes the task statement and change the progress tag of a memo to “completed.” Furthermore, users evaluate the contents of task execution. The present RAC already has self-assessment and mutual evaluation functions. However, since the memos about research activities were private, we had not previously touched upon the assessment of the contents. In this research, we enabled the RAC to unfold the memo about a task that was clarified at a seminar before laboratory members. Then, we enabled it to receive a sensitive assessment in accordance with the contents of execution of the task.
5. Experiments

The subjects that participated in the experiments ranged from undergraduate seniors (B4) to second-year graduate students (M2) in the department of information engineering and computer science. There were eight students (B4: three, M1: three, and M2: two) in our laboratory. We randomly divided each grade into the intervention group (proposed system) and the control group (conventional system) and carried out crossover comparison tests in the first semester (Apr. to Jul.) and the second semester (Oct. to Dec.) of fiscal year 2016. The comparison result for their systems is shown in Table 1.

Table 1: Evaluation of task awareness rate of proposed (P) and conventional (C) RACs.

<table>
<thead>
<tr>
<th>System</th>
<th>Ave.</th>
<th>SD</th>
<th>Effective # of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>61.4</td>
<td>23.3</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>36.2</td>
<td>32.1</td>
<td>2</td>
</tr>
</tbody>
</table>

As a result of performing a t-test of the task awareness rates between the proposed and conventional RACs, a difference was found that had the significance level of .05 (p-value = 0.0481). Moreover, when two students used the conventional system, their task awareness rates became high, but the difference between systems was several percent and thus not very large. We also administered a questionnaire to users after they had used the proposed RAC. Seven out of eight students answered “strongly agree” to the statement “Tasks I forgot were extracted.” Moreover, seven out of eight students responded positively to “RAC positively affected research activities,” “I was more motivated to organize tasks,” and “RAC is useful.”

Thus, the function for automatically extracting the task statement in the proposed system effectively distinguishes between tasks. However, in the free description of the questionnaire, users who answered “undecided” to above three statements said they had “not acquired the habit of using the system,” so we can consider using push messages to urge use of the system.

6. Conclusion

In this research, to smoothly promote a student’s research activities in a university laboratory, we added a task execution support function to a research activity concierge (RAC), which is our present research-activities support system. As a result of conducting a practical use experiment of the system for students engaged in undergraduate and postgraduate research, we found that the proposed RAC was able to grasp the existence of tasks more correctly than the present RAC.

References


AR-based Inorganic Chemistry Learning Support System using Mobile HMD

Masaru OKAMOTOa*, Tsukasa ISHIMURAa & Yukihiro MATSUBARAa
aGraduate School of Information Sciences, Hiroshima City University, Japan
*okamoto@hiroshima-cu.ac.jp

Abstract: In this paper, we proposed inorganic chemistry learning support system using smartphone. For the proposed environment, smartphone, VR viewer and some markers are used as interface. To construct smartphone based HMD, user attach smartphone to VR viewer. By wearing HMD on user’s head, user don’t need to hold smartphone with his/her hand. To perform the virtual experiment in the proposed system, a user operates marker in recorded area of smartphone’s camera. Based on recorded image, proposed system constructs virtual environment. Constructed virtual environment is displayed on smartphone’s display. By watching smartphone’s display and moving and arranging markers on the desk, user can confirm process of virtual experiments. By using proposed system, user can perform virtual experiment in various locations without PC and camera. In this paper, effectiveness of proposed approach was shown.

Keywords: inorganic chemistry, augmented reality, head mounted reality, experience based learning

1. Introduction

In chemistry of high school education, it is important that students perform experiment. However, students don’t have many opportunities to perform chemistry experiment. Okamoto et al. developed AR (augmented reality)-based learning support system for inorganic chemistry. (Okamoto et al., 2014) By using this system, a user can repeatedly perform virtual inorganic chemistry experiment. The system is constructed by USB camera, PC and markers as input interface. In the system, to perform virtual experiments, a user operates some markers that are regarded as actual instruments and solutes. A USB camera records image of the user’s operation. Based on recorded real image, the system determines the kind of marker and the position of marker. The constructed virtual environment is displayed on a display. However, it’s difficult to perform virtual experiment anywhere, because the system needs a PC and a USB camera.

Recently, many researchers developed learning support system that can be used anytime and anywhere. (Chen et al., 2012, Tabata et al., 2015, Yan et al., 2012) For constructing these systems, smartphone or a tablet device are usually used. It is easy for a learner to carry smartphone or tablet. Therefore, by using these devices, user can learn anytime and anywhere. By constructing inorganic chemistry learning support system based on smartphone, it is possible to increase learning opportunity.

In this paper, we proposed inorganic chemistry learning support system using smartphone. We construct a system using smartphone that can perform same virtual inorganic chemistry experiment as the existing system. To use same interface as the existing system, the proposed system uses some marker. In proposed system, a smartphone record marker and display virtual environment. To construct smartphone based HMD, user attach smartphone to VR viewer. By wearing HMD on user’s head and operating markers, user can perform virtual experiment using same operation as the existing system. From the above, the proposed system is constructed portable devices (markers, smartphone, VR viewer). Therefore, by using these devices, user can learn anytime and anywhere, and it is possible to increase learning opportunity.

2. Proposed Approach
2.1 System

We construct a system using smartphone that can perform same virtual inorganic chemistry experiment as the existing system. To construct the system that use same interface as existing system, the proposed system use markers, smartphone and VR viewer. Figure 1 shows a user using the proposed system. As shown in Figure 1, by wearing HMD on user’s head, user doesn’t need to hold smartphone with his/her hand. A user can operate marker using both hand same as existing system. Some markers correspond with instrument and reagent, and so on. By operating these markers in recorded area of smartphone’s camera, a user can perform virtual experiment with the proposed system.

![Figure 1. Overview of the proposed system](image1)

![Figure 2. The image of markers](image2)

Figure 2 shows an image of markers that is recorded by smartphone’s camera. In Figure 2, a beaker marker, a burner marker, and a platinum wire marker are recorded by smartphone’s camera. By recording arranged markers, a user can input information of markers into the system. Based on recorded real image, the system determines the kind of marker and the position of marker. To construct the virtual environment, the system display CGs based on user’s operation of markers. And Figure 3 shows a virtual experiment displayed on smartphone’s display. The constructed virtual environment is displayed on both side of the screen of a smartphone. The image of virtual environment is constructed based on the real image of Figure 2, and the image is displayed on the screen of the smartphone. By operating markers in recorded area, user’s operation is recognized by the system. And a user can perform various virtual experiments. In Figure 3, three instruments CGs are displayed on the real image (1: beaker, 2: gas burner, 3: platinum needle). A user can confirm the virtual environment displaying these CGs.

![Figure 3. Virtual environment displayed on the smartphone](image3)

3. Experimental Evaluation

To evaluate the effectiveness of proposed probabilistic selection of given questions and learning process using virtual environment, learning experiments about chemical reaction were conducted. Four subjects (A, B, C and D) participated in the experiments. All subject learned about the knowledge of the
chemical reaction using the proposed system. As a candidate for comparison, results of the previous systems (Okamoto et al., 2014) are used.

The results of the experiments of four subjects using proposed method are shown in Figure 4. For comparison, other four subjects using previous method are shown in Figure 5 too. All eight subjects took paper test after each experiment. By learning repeatedly, in the 17th experiment, all subjects using proposed system have acquired all the knowledge of chemical reactions corresponding to all questions given to subjects in the virtual environment. Next, to compare with a proposed system, experimental result of other subjects is shown in Figure 5. As shown in this figure, four subjects’ learning experiment is ended by 19 times too. The learning results of Subjects using previous system are like Figure 4’s result. From these results, by performing repeatedly learning using proposed system, there is possibility that a learner can get all knowledge of inorganic chemical reaction which the Japanese high school students should study.

4. Conclusion

In this paper, we proposed inorganic chemistry learning support system using smartphone. By using proposed system, user can perform virtual experiment in various locations without PC and camera. From results of evaluation experiment, it is confirmed that subjects can perform virtual experiment as directed. In future works, we would like to confirm effect of learning.

Acknowledgements

This work was partially supported by Grant-in-Aid for Scientific Research (C) No. 16K01072 from MEXT Japan.

References

Analysis on Students’ Usage of Highlighters on E-textbooks in Classroom

Yuta TANIGUCHI*, Fumiya OKUBO, Atsushi SHIMADA & Shin’ichi KONOMI

*Research Institute for Information Technology, Kyushu University, Japan
Faculty of Arts and Science, Kyushu University, Japan
Information Science and Electrical Engineering, Kyushu University, Japan
*yuta.taniguchi.y.t@gmail.com

Abstract: E-book has been gradually getting popularity in educational contexts. Reading textbooks on computers or hand-held devices enables us to track the learning activities of students regardless of situations. In our university, several courses for first year students employ our e-book system, and we have been collecting its usage logs. From the logs, it seems that the highlighter function of the e-book reader plays an important role in learning because it is used most by the students. Though many researches studied the effectiveness of e-textbooks, only limited studies addressed how students utilize highlighters and how marking activity affects their learning. In this paper, we focus on highlighted portions of e-textbooks, and analyze how students use highlighters in their learning. We also attempt to provide recommendations to students for highlighting based on the highlighter usage in other classes.

Keywords: E-book, highlighter, usage pattern, suggestion of highlighting

1. Introduction

E-book has been gradually accepted by people, and it has been also introduced to educational contexts as an alternative form of the traditional textbooks (Brown, 2013). Compared to traditional textbooks, course materials in e-book form enable us to track their usage regardless of situations. Thus, it is crucial to develop analysis methods which leads to improvement of learning by a data-driven manner.

In our university, several computer science courses use our own online e-book system in combination with the learning management system Moodle. Most classes of those courses use slides as primary teaching materials. Among the functions of our e-book reader application, highlighter feature is used most. Therefore, it is considered that the feature plays an important role in learning, and the event logs of highlighter have potential to describe the learning process of students.

The effectiveness of e-books in educational context has been studied over the years. Maynard and Cheyne (2005) studied e-textbooks in education of children, Rockinson-Szapkiw, Courduff, Carter, and Bennett (2013) made comparative study on traditional textbooks and e-textbooks, and Al-Mashaqbeh and Al Shurman (2015) studied the adoption of tablet and e-textbooks. However, how students utilize highlighter feature in their learning has not been studied well. In this paper, we investigate the usage of highlighters and discuss how such an activity affects the quiz scores.

2. Method

In this paper, we focus on the course “Primary Course of Cybersecurity” in our university. The course provides primary matters include basic technologies, laws and morals about computer security. The course is one of the mandatory courses for all first-year students. There are 15 classes for the course, and 10 teachers give lecture for around 200 students on average using the same textbooks provided through our e-book system. We chose two classes out of those classes, in which the same teacher gave lecture, as targets of highlighting suggestion. We call them “Class A” and “Class B”, for convenience.

On the left hand side of Figure 1, we show an example screen of e-book system with a material for the cybersecurity course. In the figure, a red rectangle and a yellow one are shown; students are
instructed to use red color for a portion of a page where they consider important, and yellow one for an incomprehensible part of a page. The e-book viewer records such highlightings as event logs in our database.

We chose other two classes, from which we computed a set of highlightings to recommend. These classes went ahead of the aforementioned classes by one week, and thus it was able to collect event logs in advance. We aggregated the event logs from the students of the classes, and computed the two-dimensional histograms for each page. Figure 1 shows an example heat map of accumulated highlightings on the right hand side. We extracted bounding boxes of the most notably highlighted areas that gained supports from more than 20 percent of students using highlighters at least once.

We presented obtained bounding boxes to students as highlighted areas colored blue. We provided such recommendations to Class A in the fourth week and to Class B in the fifth week. Those recommendations are limited to the course materials used in the particular weeks. We conducted quizzes at the end of every class. In the rest of the paper, we discuss the differences of the patterns of usage and acceptance of recommendations, and the difference of the scores of quizzes between these classes.

3. Results & Discussion

3.1 Usage Pattern of Highlighters

After providing recommendations in classes, we investigated the event logs of highlighters to see how students responded to the recommendations. Figure 2 shows a heat map table of event frequencies for Class A in the fourth week, and Figure 3 shows the same for Class B in the fifth week. Since the tables are very wide, they are split into two lines. Columns correspond to students, and rows correspond to the types of highlighted regions: the numbers of remaining suggested highlightings (first row), the numbers of added highlightings (second row), and the numbers of deleted highlightings (third row). Every cell is colored according to its count value, where the highest value corresponds to red and the lowest value corresponds to green.

Figure 2. A heat map table of counts of highlighter-related events for Class A.

Figure 3. A heat map table of counts of highlighter-related events for Class B.
The upper matrices shows counts especially for students who did not use any highlighters, and the lower matrices shows the counts especially for students who used some highlighters. From the upper matrices of the both figures, we can see that most of the students did not care about the suggested highlightings. In contrast to that, from the lower matrices, more students made reaction to the suggested highlightings in addition to their own highlightings. We can say that whether a student is concerned about highlighters or not is strongly related to whether the student care about the suggestion or not.

3.2 Highlighting Suggestions and Scores of Quizzes

Table 1 shows the average scores of quizzes for each class on fourth and fifth weeks. Asterisks beside numbers represent that students are given highlighting suggestions. From the result, we cannot say that suggesting highlightings does or does not improve the quiz scores.

Table 1: Average scores of quizzes.

<table>
<thead>
<tr>
<th></th>
<th>Class A</th>
<th>Class B</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th Week</td>
<td>6.21*</td>
<td>6.35</td>
</tr>
<tr>
<td>5th Week</td>
<td>6.10</td>
<td>7.94*</td>
</tr>
</tbody>
</table>

However, compared to the fourth week, the average score of students of Class B in fifth week is relatively high. We further investigated whether suggested highlightings are related to the content of quizzes or not, and we found that suggestions given to Class B in the fifth week is related to quizzes more than those given to Class B in the fourth week. Therefore, we can conclude that just suggesting the most popular areas does not necessarily improve scores although there is some possibility that considering the content of highlighted area could give a good influence to students.

4. Conclusion

We showed a preliminary result of analysis on the usage of highlighters for e-books. We discussed how our recommendation of highlighters affect students’ learning. There were almost the same numbers of students who uses highlighters and who do not use them, and their responses to the recommendations are somewhat different. Although we could not find any evidence that suggestions themselves improve students’ quiz scores, it was indicated that we may be able to see some improvement in quiz scores if we examine the contents under the highlightings in addition to the popularity of the areas.

Acknowledgements

These research results have been achieved under the theme of “Research and Development on Fundamental and Utilization Technologies for Social Big Data” (178A03), as Commissioned Research for the National Institute of Information and Communications Technology (NICT), Japan.

References


A Taxonomy for Teacher-Actionable Insights in Learning Analytics

Elizabeth KOH* & Jennifer Pei-Ling TAN
National Institute of Education, Nanyang Technological University, Singapore
*elizabeth.koh@nie.edu.sg

Abstract: In the field of learning analytics (LA), actionable insight from LA designs tends to be a buzzword without clear understandings. As the teacher is a key stakeholder, what teacher-actionable insights can be derived from LA designs? Towards providing greater clarity on this issue, we concretize a taxonomy of LA decision support for teacher-actionable insights in student engagement. Four types of decision support are conceived in this taxonomy with relevant teacher implications. Through this taxonomy, we hope to offer possible pathways for actionable insight in LA designs and make clearer the role of the teacher.

Keywords: Learning analytics, teacher, design, taxonomy, actionable insight

1. Introduction

In the field of data analytics, the term “actionable insight” is a buzzword without clear definitions. Recognizing this, Tan and Chan (2015) provide a three-tiered definition for actionable insights in data analytics systems: analytic insight (understanding and inferring individual information), synergistic insight (contextualizing, combining and linking information), and prognostic insight (deriving information of future results). Similarly, others have highlighted an analytics continuum ranging from descriptive, diagnostic, to predictive and prescriptive analytics (Gartner.com). In the field of Learning Analytics (LA), concomitantly, several conceptions and understandings of actionable insight exist from different stakeholders such as learners and teachers (Lu, Huang, Huang, & Yang, 2017; Sergis & Sampson, 2017). Evident from extant literature is that “insight” can be understood in several ways and from different stakeholders. Many LA designs have focused on providing interventions such as tasks and recommendations for the learner. However, comparably less attention is paid to a closely intertwined stakeholder, the teacher (Sergis & Sampson, 2017). Especially within the K-12 education context, where the teacher more often than not plays a crucial ‘make-or-break’ role in the learning and teaching process (Hattie, Masters, & Birch, 2015), teacher-actionable insights are important.

What teacher-actionable insights can be derived from LA systems? Towards scoping this question, we premise the design of many LA systems in the area of engagement in learning. In the pedagogical core of learning there is an interaction between learners and the content, as well as between peer learners (Tan & Koh, 2017). Hence, student engagement is commonly measured in LA designs and used to inform actionable insight such as through the engagement of students with the content, and with other peers (Lu et al., 2017; Tan & Koh, 2017).

We posit that LA can provide teacher-actionable insights for understanding this engagement in learning. As such, we conceptualize a taxonomy of LA decision support for teacher-actionable insights in student engagement. This taxonomy is briefly illustrated using the Collaborative Video Annotation and LA (CoVAA) Learning Environment, a prototype LA time-point based video annotation system.

2. Conceptualizing a Taxonomy for Teacher-Actionable Insights in LA Designs

Informed by the literature, we conceptualize four types of LA decision support for teacher-actionable insights in student engagement: descriptive, diagnostic, predictive and prescriptive (Table 1). The second column in Table 1 describes the areas of teacher-actionable insight, which is a more macro view...
of system feedback to the teacher. The third column highlights certain data science methodologies and techniques required while the last column provides implications of this decision support for teachers.

Table 1: A taxonomy of LA decision support for teacher-actionable insights in student engagement

<table>
<thead>
<tr>
<th>Type of LA decision support</th>
<th>Areas of teacher-actionable insights</th>
<th>Possible data science methodologies</th>
<th>Implications for teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive</td>
<td>What are students engaged in? What are they doing, feeling, learning?</td>
<td>Dashboard summaries, visualizations, descriptive statistics</td>
<td>Broad ranging areas of action, relies on the agency of teachers</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>Why are students’ engaged?</td>
<td>Visualizations, process mining, drill-down tools, correlations, data discovery, and data mining</td>
<td>More specific areas of action, but still requires teacher discernment for intervention</td>
</tr>
<tr>
<td>Predictive</td>
<td>What will students’ be engaged in? Which groups of students’ will be engaged?</td>
<td>Machine learning, regression analysis</td>
<td>Relieves load of teachers for certain areas of action, but could provide other opportunities for teachers to look at other areas of engagement</td>
</tr>
<tr>
<td>Prescriptive</td>
<td>What can be done to engage students?</td>
<td>Machine learning, algorithms, predefined conditions</td>
<td></td>
</tr>
</tbody>
</table>

2.1 Descriptive

Descriptive analytics describes what students’ activities on the system are, depicting indicators of student engagement for the teacher. It represents the foundational data structures in LA and describes what students’ activities on the system are. For instance, in CoVAA, teachers are able to download a set of participation data including annotation type, critical lens tag, and comment description, which makes it convenient for them to examine and provide feedback on students’ answers. Many LA designs provide such engagement data in real-time so teachers are able to see and monitor the activities of students instantaneously. Descriptive analytics typically summarize different engagement types for teachers using descriptive statistics in graphs etc. Still challenges of what metrics to measure as learning designs become more sophisticated and how best to represent them exist.

Teacher-actionable insight at this layer tends to directly relate to the metric or indicator measured e.g., submission data. Besides giving the teacher an aggregated understanding of the students, and/or comparison of learners, the LA engine typically does not provide further decision support for the teacher. Teacher actionable insight depends on the capacity and agency of the teacher to take action. Teachers have to make sense of the data and decide for themselves appropriate interventions (Sergis & Sampson, 2017). In that sense, descriptive analytics offers broad ranging areas of teacher-actionable insights, but also relies on the capacity of teachers to decide and perform more targeted interventions.

2.2 Diagnostic

Diagnostic analytics tries to explain why students did what they did using data science methodologies and techniques including visualizations, process mining, correlations, data discovery, and data mining.

This LA design attempts to link relationships to explain student engagement and helps teachers to pinpoint specific areas for possible interventions. Still, teachers should be discerning and decide pedagogically if they should intervene. For CoVAA, this layer of diagnosis is currently done in the back-end using existing statistical techniques by researchers, and shared with the teachers, as data-driven evidence for teachers to take action.

2.3 Predictive and prescriptive
Predictive and prescriptive analytics are closely related. While predictive analytics provide empirical evidence of what students will be engaged in, prescriptive analytics provide recommendations to the student, reducing the immediate intervention required by the teacher. Predictive analytics provide empirical evidence of what students will be engaged in. This layer provides teachers with foresight, what will happen based on probability estimates. On the other hand, prescriptive analytics asks “what can be done to engage students” and prescribes actions that the system takes on behalf of the teacher. It computes activities and responses that the system can do now based on predefined conditions that were determined by diagnostic and predictive analyses.

Predictive analytics provides very clear and specific teacher-actionable insight. Decision support for the teacher is precise and could include filtering and identifying different clusters of students such as those potentially at risk from academic failure and dropout. It can also identify students who are potentially on an accelerated trajectory. Teachers’ usage of system tools can also predict student achievement. Prescriptive analytics then seeks to identify specific sets of activities that students can take, without the immediate intervention from the teacher.

While on one hand these two types of support may seem to reduce the need for the teacher, we posit that at the same time, this provides opportunities for teachers to go beyond the common set of responses to probe deeper into student engagement or examine new trends among their students. Seemingly, this helps relieve the load of teachers’ direct instruction to the student, and could help the teacher to focus on other areas of student engagement that is not provided for by the system. As such solutions require more time and testing, these analytics are part of the future work planned in CoVAA.

3. Concluding Remarks

We have conceptualized a taxonomy of LA decision support for teacher-actionable insights in student engagement comprising four types. While these types may seem to have some sort hierarchical relationship, e.g., each type being a more complex type of the other, we realize that each could uncover engagement ranging from the superficial, simple to complex and deep. Each type then offers various pathways of providing feedback to teachers. Teacher-actionable insights in student engagement is a crucial area for the emerging field of LA, and in clarifying possible pathways, LA designs can be made more useful for teaching and learning.

Acknowledgements

This paper refers to data and analysis from the project NRF2015-EDU001-IHL08 and NRF2015-EDU001-IHL09 funded by the Singapore National Research Foundation, eduLab Research Program. The views expressed in this paper are the authors’ and do not necessarily represent the views of the National Institute of Education.

References

System Architecture and Predictive Experiment for an Automatic Learning Support Function on Classroom Response Systems

Kozo Mizutani*
Department of Information and Electronic Engineering, School of Science and Engineering, Teikyo University, Japan.
*mizutani@ics.teikyo-u.ac.jp

Abstract: The Classroom Response System (CRS) was developed as a tool for enhancing interactivity between teachers and students in the classroom. Many studies have revealed the effectiveness of using CRS in educational environments. However, the popularity of CRS has not increased sufficiently. This study aims to implement an automatic learning support function on CRS to improve conventional systems. In this regard, I considered applying the concept of an agent-oriented system (AOS) to realize the implementation of this support function on web-based CRS. In addition, I examined the feasibility of applying the concept of an automatic learning support function by examining the prediction of student responses by using student activity logs. This paper describes the model on which the application of an AOS is based and the result of the prediction examination.

Keywords: Classroom Response System, Clicker, Learning Analytics, Multi-Agent System

1. Introduction

The Classroom Response System (CRS), also known as a Student Response System (SRS), Audience Response System (ARS), or Clicker, is a tool for enhancing interactivity between teachers and students in the classroom. The use of CRS has become accepted as a pedagogical tool capable of increasing the effectiveness of education. However, although the educational effect is appreciated, CRS has not become sufficiently popular. One of the reasons is considered to be that learning support functions that adapt to each student are not implemented on CRS or that its effect is insufficient. Therefore, my approach has involved considering methods to realize a learning support function that enables automatic and real-time improvement of further education using CRS. This paper describes the system architecture for realizing the automatic learning support function and the results of the prediction examination of student responses using activity logs based on the architecture.

2. Background

2.1 Related Work

Studies of the system architecture of CRS and the pedagogy of using the system in practice are important themes of information technology in education and have been attracting attention for the past few decades (Fujita et al. 1969). Recently, Karakostas et al. developed a CRS named "QuizIt" for Android devices (Karakostas et al. 2014). QuizIt consists of three parts: a mobile application for the lecturer, a mobile application for the students, and a web environment for lecturer administration. All of these components run in real time. The results of a pilot study showed that students were quite positive about the use of QuizIt in the laboratory course. Barth-Cohen et al. investigated the effect of promoting productive peer discussion among middle school science students using Clicker (Barth-Cohen et al. 2016). Their study recorded peer conversations to characterize the nature of student discourses and an analysis of these conversations. In another study using Clicker to investigate peer discussion, Majumdar
et al. tried to visualize and categorize Clicker responses using the interactive stratified attribute tracking diagram (iSAT) (Majumdar et al. 2015). They concluded that iSAT contributes to conducting cohort analysis of teaching learning practice for teachers. The result revealed that students improve their performance on Clicker questions when they engage in peer discussion. Premkumar used Clicker in summative exams (Premkumar 2016). The results showed the process of an instructor is less time consuming, efficient, and more secure compared to the use of scan sheets. Furthermore, students were accepting the use of this technology in high stakes exams, and found it engaging and satisfying, primarily because it provides them with instant feedback. In addition, many studies about using CRS relating to the effectiveness of feedback and timing (Lantz et al. 2014), the effect of different teaching strategies (Liu et al. 2016), and so on, have been reported.

In addition to these studies about the use of CRS, other studies relating to evaluating the impact of using the CRS have also been published. For example, Hunsu et al. examined the potential effects of using CRS by comparing classrooms that did and did not use it to determine the different cognitive and non-cognitive learning outcomes (Hunsu et al. 2016). They discovered that CRS has a small but significant effect on cognitive learning outcomes and a near medium effect on non-cognitive learning outcomes. They also pointed out that instructors need to provide feedback in a constructive and timely manner. Richardson et al. mentioned the necessity of a standardized instrument to evaluate the impact of using the CRS, and developed an instrument named the CRiSP questionnaire (Richardson 2015). The questionnaire consists of 26 base items and includes three scales: the usability, the impact of using the CRS on engagement, and the impact of using the CRS on learning.

In summary, the aforementioned previous studies evaluating the impact of using CRS show that, although the educational effect is appreciated, CRS has not become sufficiently popular. Many studies have investigated the educational effect of using CRS; however, reducing the operational costs associated with CRS is important for its popularization. CRiSP does not include items of system operation. Moreover, we need even more discussion about the realization of automatic learning functions in which the system grasps the state of student understanding, presents additional teaching materials, or warns irresponsible students. Therefore, the learning effect of using CRS is expected to be increased by the introduction of automatic learning support functions.

2.2 Approach of This Study

Based on the above-mentioned background, this study developed two different CRSs named "Response Analyst" that runs on PCs and "WebCRS" that runs on smart devices. These CRSs have been used in actual classes (Mizutani 2013, 2014) and aim to realize automatic learning support functions by applying the experience that was gained during past use of these systems.

These systems have the following characteristics:

- **Real-time communication:**
  Because CRS helps communication in classes, real time is important. Response Analyst uses the original protocols to optimize the real-time response of the system. WebCRS relies on the use of WebSocket, which is a new web technology for real-time communication.

- **Collect and Record Usage Logs:**
  The systems have a function that collects and records usage logs in the form of an "Activity Log" with millisecond precision. On the students’ devices, WebCRS not only makes it possible to transmit data relating to the basic functions such as receiving question contents, receiving a signal to start answering, and sending an answer, but also enables data to be collected by the accelerator sensor or inclination sensor of a smart device (Mizutani 2016). These data are recorded to a database (DB) in the background with the aim of reducing the influence on the basic functions of CRS by maintaining a small system load. In addition, the log data of the basic functions are collected with a timestamp of both the servers and each student's device. The timestamp of a student’s device can be used to accurately analyze the state of a student. For example, the time that passes from the moment a student receives a start answering signal until the moment they submit their answer cannot be accurately analyzed by using the server time because of the time error caused by network delay and so on. By using a device timestamp, an accurate analysis of the difference of the time of each event is possible.
Moreover, WebCRS has the following characteristics, designed to improve the problems mentioned above in section 2.1.

- Using smart devices:
  WebCRS supports common web browsers pre-installed on PCs, smartphones, tablets, and so on.

- No installation required:
  To be able to start using WebCRS easily, WebCRS is implemented as a web application. If a web browser is installed on a device, we can use WebCRS without the installation of other programs. In case communication by WebSocket cannot be established by limitation of the network, browser version, and so on, WebCRS uses the conventional HTTP-based protocol. As a result, using WebCRS as a “bring your own device” (BYOD) is possible too.

- Reducing Costs:
  The costs of using the entire system from preparation to maintenance after classes are reduced. By adopting BYOD, the cost of managing the student's device can be avoided. The Web servers, as the system core of WebCRS, also run on public cloud services.

- Scalability:
  The architecture of WebCRS is designed for cloud services; thus, it is easy to multiplex web servers. The design assumes simultaneous use at a scale ranging from one classroom to an entire school.

Figure 1 shows an example of a screenshot of the WebCRS interface. The functions for students and teachers are provided as web applications.

![Figure 1. Examples of WebCRS screens.](image)

The next section describes concrete methods intended to achieve the implementation of an automatic learning support function on WebCRS.

### 3. System Design for an Automatic Learning Support Function

#### 3.1 An Automatic Learning Support Function

This study aims to realize an automatic learning support function as follows by using the activity logs of students on WebCRS.

- Providing supplementary teaching materials:
  To help a student to answer a question, the system decides whether to provide the student with supplementary teaching materials, and automatically provides these materials if the system considers this necessary.

- Warning to irresponsible students:
  The system detects irresponsible students who answer a question without sufficient thinking, and warns the students to learn positively.
• Notification to a teacher:
The system notifies a teacher of the state of these supports.

In this section, concrete methods to implement these functions on WebCRS are considered.

3.2 Problems Associated with Implementing the Learning Support Function

The architecture of WebCRS, which is designed to enhance scalability, is shown in Figure 2. The system includes a conventional web application, web servers, and DB servers, but a load balancer (LB) is introduced to distribute the requests for user connections.

The web server includes a mechanism implemented to process distributed requests by the LB. For example, as shown in Figure 3 (a), the request of Student A is connected to Node A by the LB. Node A stores the state of the connection as Session Info to the DB. Thereafter, the load balance of the system is changed, and the destination of the connection of Student A is changed to Node B by the LB, as shown in Figure 3 (b). In this case, because Node B does not have the Session Info of Student A, Node B obtains this information from the DB and maintains the state of the connection continuously.

Figure 2. System structure of WebCRS.

Figure 3. Mechanism of Session Info on WebCRS.

Session Info is always stored to the DB to enable current Session Info to be shared by each web server node even when the LB changes connections. As a result, verification of the scalability of WebCRS confirmed that the effect of load distribution is reduced when the system is used simultaneously by over 400 students (Mizutani, 2015). An investigation attributed this decline in performance to the load caused by recording Session Info to the DB. When using the Activity Log for the learning support function, the load of the DB is increased, thereby increasing the possibility of influencing the basic functions of the CRS. To realize the learning support function, a mechanism for using the Activity Log capable of maintaining system scalability is needed.
3.3 Applying the Concept of an Agent-Oriented System

This study attempted to resolve the above-mentioned problem by applying the concept of an Agent-Oriented System (AOS) to the system architecture of WebCRS. The AOS concept employs a Multi-Agent, which is an aggregation of an agent as an element of the system. An important concept of the AOS is a Mobile Agent, which is an agent capable of moving on networking, and is defined in FIPA as a standard specification of AOS (FIPA 98).

Figure 4 shows an example of system behavior adopting a Mobile Agent. The Multi-Agent layer is cross-sectionally employed in each node of the web servers. In each node, an agent that processes Session Info is created each time a student connection is established. The agent is a Mobile Agent that is able to move on the Multi-Agent layer; thus, if the destination of the connection is changed, the agent moves to a new node. Because it is able to move by retaining the Session Info, it becomes unnecessary to obtain Session Info from the DB. Furthermore, it also becomes unnecessary to always store Session Info to the DB, thereby reducing the load of the DB.

3.4 Prediction Using Activity Log for the Learning Support Function

To realize the automatic learning support function, three decisions need to be made: timing of support, a decision whether to provide support, and a method of support. For example, in relation to the timing of support, a function to provide supplementary teaching materials support includes a method to provide timing when a student answer is incorrect. However, this method may allow a student who answers randomly without thinking about a question to unintentionally receive supplementary teaching material.

As to whether to decide to provide support, it is thought to provide the materials when the average of the correct answer of a student is less than a threshold that is configured beforehand. However, because the difficulty of each question varies, it is not possible to effectively decide by using a simple threshold only. Ideally, the educational effect would be enhanced by providing materials before a student submits their answer only when the possibility exists that their answer is incorrect.

In addition, it is necessary to realize this by reducing the load of the DB. Consequently, this study tries to predict whether the $N$th answer of a student is correct or incorrect by using the Activity Log of until the $N-1$th answer, assuming to use the log maintained by the agents.

In this prediction, the following items in the Activity Log are used:

- Correct / Incorrect value of a response ($Cr$)
- Response Time required for each response ($Rt$)
- Label of limit time of each question ($Limit$)
- Activate Time for an response ($Act$)
- Flag value that means the time of a response is shorter or longer than the average of all students. ($ActF$)
Figure 5 illustrates these items. $R_t$ is the time between the moment a student's device received the message to start answering and the moment the student answered. Because of the time limit for answering each question, it is normalized into the range $[0, 1]$. $\text{Limit}$ is a label that signifies the actual length of the time limit for answering as indicated in Table 1.

![Diagram](image)

Figure 5. The items of Activity Log in the prediction.

Table 1: Labels indicating the Time Limit for answering.

<table>
<thead>
<tr>
<th>Labels of $\text{Limit}$</th>
<th>Actual time limit for answering (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$t \leq 20$</td>
</tr>
<tr>
<td>B</td>
<td>$20 &lt; t \leq 40$</td>
</tr>
<tr>
<td>C</td>
<td>$40 &lt; t \leq 60$</td>
</tr>
<tr>
<td>D</td>
<td>$60 &lt; t \leq 120$</td>
</tr>
<tr>
<td>E</td>
<td>$120 &lt; t$</td>
</tr>
</tbody>
</table>

$\text{Act}$ is the time between the moment the answer started and the time at which the CRS screen is activated by a student. Irresponsible students may sometimes be shown a screen other than the CRS. These students activate the screen of the CRS after they start answering and answer a question immediately after activation without thinking about the question. Because such student behavior might influence the correctness of an answer, $\text{Act}$ is included in prediction data. Similar to $R_t$, $\text{Act}$ uses the value that is normalized by the time limit for answering each question.

The purpose of the prediction based on the prediction result, depending on whether the $N$th answer of a student is correct or incorrect, it is to provide materials before the $N$th answering event only when the possibility exists that the student answer is incorrect. At the moment just before the $N$th answering event of a student, it is not possible to know whether the $N$th answering time of the student is faster or slower than the average of the $N$th answering time of all students. In this case, $\text{ActF}$ uses the values until the $N-1$th only for prediction.

The Activity Log of $N$ times of answering forms one set of data, $\text{DataSet}_i$, which is defined as follows.

$$\text{DataSet}_i = \{C_r, ..., C_{r_N}, R_t, ..., R_{t_N}, \text{Limit}_1, ..., \text{Limit}_N, \text{Act}_1, ..., \text{Act}_N, \text{ActF}_1, ..., \text{ActF}_{N-1}\}$$

Here, $i=1, 2, ..., I$, $I$ is the number of datasets that can possibly be used for prediction. For example, when the number of all answering events that are stored as the Activity Log is 6, if $N=5$, it becomes $I=2$. That is, the aim of this study is to predict $C_{r_N}$.

Furthermore, $C_r$, $R_t$, $\text{Limit}$, and $\text{Act}$ are data that the agent is able to hold without access to the DB on the system model that applies the AOS. Because $\text{ActF}$ is a flag to indicate the average answering time of all students, it is necessary to calculate the answering time that is held by each agent. This can be implemented by using the Blackboard Model of AOS.
4. Prediction Experiment

The above-mentioned items are predicted by using the Activity Log that consists of actual data in classrooms that introduced CRS. The Activity Log of this study consists of data obtained by using the "Response Analyst" in classes of students who attended an information technology course in our university. The course syllabus includes solving past versions of the Information Technology Passport Examination (IT Passport Exam), one of the Information Technology Engineers Examinations in Japan. The IT Passport Exam consists of multiple-choice questions. Examinees select one answer from among four choices in response to each question.

A total of 1,863 Activity Logs were stored. This included the logs of 53 students who attended the course, 50 questions from past versions of the IT Passport Exam, and the data concerned with this usage. Using these Activity Logs, the dataset for prediction is created. Three datasets are created by changing the value of $N$. The number of datasets $I$ in each $N$ is shown Table 2. These datasets are divided randomly according to a ratio of 3:1 for two purposes: (a) for training a classifier, (b) for testing the classifier.

The target of prediction is a correct or incorrect $N$th answer. This study uses two classifiers for the prediction: (1) A two-class neural network that has three fully connected hidden layers, (2) a two-class locally deep support vector machine (Deep SVN, Jose 2013). The neural network model of (1) is shown in Figure 6 and the result of the prediction is shown in Table 3.

Accuracy is the rate of correctly predicted results to the testing data. These are including the cases a student answer becomes correct and the cases it becomes incorrect. Precision is the rate of correctly predicted results to the predicted results that are a student answer becomes correct. Recall is the rate of correctly predicted to actual correct answers in the testing data. F-Score (F-measure) and AUC are indicators using these values, when these become to be 1.0, it represents the prediction result is good.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>676</td>
</tr>
<tr>
<td>4</td>
<td>523</td>
</tr>
<tr>
<td>5</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 2: Number of datasets.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>$N$</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep SVN</td>
<td>3</td>
<td>0.568</td>
<td>0.560</td>
<td>0.659</td>
<td>0.605</td>
<td>0.568</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.550</td>
<td>0.566</td>
<td>0.671</td>
<td>0.614</td>
<td>0.552</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.580</td>
<td>0.608</td>
<td>0.585</td>
<td>0.596</td>
<td>0.561</td>
</tr>
<tr>
<td>Neural Network (NN)</td>
<td>3</td>
<td>0.497</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.560</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.649</td>
<td>0.633</td>
<td>0.814</td>
<td>0.713</td>
<td>0.640</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.510</td>
<td>0.531</td>
<td>0.642</td>
<td>0.581</td>
<td>0.520</td>
</tr>
</tbody>
</table>

Figure 6. Definition of neural network for prediction, written in Net#.

5. Consideration

To realize an automatic learning support function, this study considered applying the concept of AOS to WebCRS. It is thought that the effect of load balancing is improved and the load of the DB system is reduced. As a result, it is possible to realize automatic learning support functions using the Activity Log.
of students stored in the DB. On other hand, the compatibility between an agent framework and the web application framework used by WebCRS presents a problem. In the current state, WebCRS is able to operate on SaaS type public cloud services. Major SaaS frameworks do not support the implementation of agent frameworks such as JADE (Bellifemine, 1999). Although this problem can be resolved by the implementation of unique mechanisms, the compatibility might decrease. Therefore, it is necessary to consider a method for the implementation of an agent framework to maintain compatibility with cloud services.

In terms of the experimental results for the correct / incorrect prediction of answers using an Activity Log, the best result was the case of $N=4$ when using a neural network (NN). It shows that it is able to predict the answer of a student, with a probability of 0.649. When $N=3$, the NN resulted in overfitting. When $N=5$, the result decreased in comparison to the case of $N=4$. The reason for this is thought to be that the number of datasets is smaller than the case of $N=4$. In addition, this study did not classify the dataset according to each student. Because the time required for providing a correct answer and the learning attitude are different for each student, the prediction results are expected to improve when the dataset is classified according to each student. However, this would increase the size of the Activity Log of each student.

The probability of an answer being correct or incorrect is 0.5. The Activity Log used in this experiment involves responses that consist of one answer from among four choices in a question; thus, the probability of a correct answer is 0.25. And also, the actual ratio of correct answers of the all students is 0.523. Each item of the Activity Log using the prediction does not have a direct logical relation to each other. However, when using CRS, it is possible to indirectly interpret the extent to which a student understands the learning material as the time they require to answer, their behavior in terms of switching screens to and from other application, and so on. Because the probability of the results including DeepSVN is over 0.5, it is possible to predict whether an answer is correct beforehand. By improving the accuracy of the prediction, it is thought that a function providing supplementary teaching materials before the student answers when an incorrect answer is predicted can be realized.

6. Conclusion

This paper discussed the problems of conventional CRS, described plans for resolving the problems and to implement an automatic learning support function. On web-based CRS, a learning support function that adapts to the state of student answering could be realized by applying the concept of AOS as system architecture. Effective learning support was achieved by attempting to predict a student answer by applying the architecture.

By applying AOS, it is possible to realize automatic learning functions while maintaining the scalability and usability as a basic function of CRS. Although the accuracy of the prediction results was insufficient, the possibility of realizing an automatic learning support function that predicts whether a student answer is correct or incorrect before they answer, and automatically provides supplementary teaching materials was described.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP15K16262.

References


Enhancing Seamless Learning Using Learning Log System

Noriko UOSAKI*, Hiroaki OGATAb, Kousuke MOURIc & Chengjiu YINd

aCenter for International Education and Exchange, Osaka University, Japan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan
cDepartment of Computer and Information Science, Tokyo University of Agriculture and Technology, Japan
dInformation Science and Technology Center, Kobe University, Japan

* n.uosaki@gmail.com

Abstract: In this paper, we describe implementation of seamless language learning with a learning log system called SCROLL and an e-book system. Prevalence of high-performance mobile devices has enhanced the potential of seamless language learning environment. Seamless learning is an approach which enables learners to entwine various learning scenarios seamlessly. We have seen a good deal of technology-driven researches in this field. Our SCROLL project is among them. SCROLL is expected to facilitate learners to learn their target languages and language related knowledge in a seamless learning environment. The evaluation was conducted where SCROLL was introduced in Japanese language learning settings. The result of the evaluations revealed that the whole system had a high usability, and it helped users gain related knowledge seamlessly.

Keywords: Learning log, mobile assisted language learning, Japanese language learning, seamless learning

1. Introduction

According to Article 21 of Standards for Establishment of Universities, 45 hours are required to provide one credit. It is a common practice for Japanese universities to give 2 credits by taking 1.5-hour-class for 15 times for one semester. In this case inside-class learning time is 22.5 hours. It is apparent that inside-class learning time is far from sufficient. Therefore it is highly necessary to promote out-of-class learning.

Especially out-of-class learning is pivotal as for foreign language learning. According to the Foreign Service Institute (FSI) of the US Department of State, 2,200 class hours are necessary for English speakers to achieve general professional proficiency level of Japanese language. Obviously it is impossible to run class for 2,200 hours at any level of school education. Lack of learning time is one of the serious problems in the second language learning education. Seamless learning environment, which enables smooth transitions between any learning scenarios such as formal and informal; personalized and social learning; learning beyond time and space, is expected to enhance their out-of-class learning.

2. Related Research

2.1 Seamless Learning

‘Seamless learning’ describes the situations where students can learn whenever they want to in a variety of scenarios and that they can switch from one scenario to another easily and quickly using one device or more per student (‘one-to-one’) as a mediator (Chan et al., 2006). Wong and Looi (2011) identified ten salient features of seamless learning; (1) Encompassing formal and informal learning, (2) Encompassing personalized and social learning; (3) Across time, (4) Across locations, (5) Ubiquitous...
knowledge access, (6) Encompassing physical and digital worlds, (7) Combined use of multiple device
types, (8) Seamless switching between multiple learning tasks, (9) Knowledge synthesis, (10)
Encompassing multiple pedagogical or learning activity models. More than 20 years ago when ‘seamless
learning’ was hardly a well-known term in the field of pedagogy, American College
Personnel Association (1994) stressed the importance of linking students’ in-class and out-of-class
experiences to create seamless learning and academic success (Wong and Looi, 2011).

Since seamless learning proposed, many researchers explored their seamless learning based
projects. Milrad et al. (2013) introduced five different seamless learning projects researched in Taiwan,
United Kingdom, Sweden, Singapore, and Japan. Uosaki et al. (2010) proposed a seamless learning
system called SMALL System (Seamless Mobile-Assisted Language Learning Support System).
Meanwhile, Wong et al. (2014) proposed a system called MyCloud (My Chinese Language ubiquitOUs
learning Days), where they explored the integration of mobile and cloud technologies for self-directed,
collaborative and seamless Chinese Language learning among primary students.

2.2 Mobile Assisted Language Learning

Mobile technologies have been expected to foster shifting from classroom-based learning to the one
that is free from time and space boundaries. Therefore they play a critical role in implementation of
seamless learning. Since our target domain is language learning, our research is closely related to mobile
assisted language learning (MALL). Mobile technologies open the door for a new kind of learning
called ‘here and now learning’ that occurs when learners have access to information anytime and
anywhere to perform authentic activities in the context of their learning (Martin & Ertzberger, 2013). It
is often argued that mobile devices are particularly suited to supporting social interaction and
collaborative learning - claims that have obvious relevance for language learning (Kukulska-Hulme &
Shield, 2008). Viberg & Grönlund (2012) reviewed the literature on MALL research from 2007 to 2012
and reported that most studies supported the hypothesis that mobile technology could enhance learners’
second language acquisition.

Ogata et al. (2004) proposed TANGO (Tag Added learNinG Objects) system which employed
the physical objects using RFID tags for language learning. Stockwell (2007) developed a prototype of
mobile-based intelligent vocabulary learning system called Vocab Tutor. Chen and Chung (2008)
developed personalized mobile English vocabulary learning system based on Item Response Theory
and learning memory cycle. Li et al. (2010) evolved an adaptive Kanji learning system using mobile
phones. Underwood et al. (2010) developed a mobile-based self-initiated vocabulary learning
application called m-iLexicon.

We have witnessed a growing presence of smartphones, tablets and other mobile devices for
more than a decade. MALL has been recognized as one of the natural directions toward which CALL
(Computer-Assisted Language Learning) is heading (Chinnery, 2006; Stockwell, 2007). Thornton and
Houser (2005), who reported that the learners preferred mobile platforms over PCs, also endorsed this
trend. However, there are some negative aspects reported in mobile learning. Mobility of learning was
not necessarily synonymous with its unlimited flexibility or learning ‘anytime anywhere;’ neither was
unlimited flexibility desirable in all learning situations and educational contexts (Palalas, 2015). There
was often a complaint from users about the user-unfriendliness of small screens on mobile devices
(Uosaki & Ogata, 2009). The small screen issue is what we still need to consider.

3. SCROLL

Learning can happen anytime, anywhere. When we come across new knowledge, people may take
notes. However, the notes will not remind us of what we have learned. SCROLL has been developed in
order to support learners to record what they have learned in both informal and formal setting using a
web browser and mobile device as a log and to share them with other learners anytime and anywhere
beyond space seamlessly, and link their past learning with their future learning (Ogata et al. 2011).
SCROLL is a client-server application. The server side runs on Linux OS. It runs on different platforms
such as smart phones, PCs, and tablets. This on-going project is still in progress with new functions
being added to the system one after the other. Our latest addition is an e-book system. The past studies
show that the system is effective for learners with their vocabulary learning, it contributes to linking in-class learning with outside of class learning, to boosting outside-of-class learning and to enhancing learners’ learning opportunities (Uosaki et al, 2012; Uosaki et al. 2014, Mouri et al, 2015). Fig. 1 shows its log-in and home interface on mobile. Its functions are designed to support implementation of seamless learning environment.

![Figure 1. SCROLL Log in Interface (right) and its Home Interface (left).](image)

### 3.1 Functions

#### 3.1.1 Recording

It facilitates the way learners record their newly learned terms to the server. For example, when a learner comes across a new term, ‘御社’ (honorific language meaning ‘your company’, one of typical business Japanese) while he was reading job-search related contents (Fig. 2 left), he can upload it to the system with texts, images, video, or pdf files. Translation is facilitated by Google translator (Fig. 2 right).

![Figure 2. New word, ‘御社’ while Reading (left) and SCROLL Supports Logging (right).](image)
3.1.2 Recommendation

SCROLL recommendation function goes as follows: If he/she uploaded ‘御社’ to the system, the system checks if the same log or related logs were already uploaded or not and show them as the related terms (Fig. 3). Human beings are likely to forget. Even though he thought it was new to him, he might have already learned it before in the past and uploaded it to the system already. Then the system links his new log with his past log. This recommendation function powerfully assist the implementation of seamless learning: to link e-textbook learning with learning through real life experience, to link learners’ present learning with their past learning and to link a learner’s learning with other learners’ learning.

3.1.3 Re-logging

Re-logging function assists to link one learner with other learners beyond time and space seamlessly. When a learner sees other learners’ log and find it useful, he/she can ‘re-log’ it to make it his/her own log just like ‘retweet’ in Twitter. For instance, if they want to learn ‘B to C’, which was uploaded by someone else, they click ‘Click to re-log’ button as shown in Fig. 4. Then it appears in their ‘My Logs’ page. Therefore, learners can obtain knowledge from others without having experienced it themselves. Using this function, knowledge can be shared by users seamlessly beyond time and space.

![Figure 3. Related terms.](image1)  ![Figure 4. ‘Re-log’ button of SCROLL.](image2)

3.1.4 Quizzes

It is reported that quiz function is effective to reinforce their memory (Li et al. 2013; Uosaki et al. 2013). The quiz function also assists to link their present learning with their past learning. Four types of quizzes (combination of image and text, multiple-choice and yes-no quiz) are generated automatically by the system. Fig. 5 shows interfaces of a multiple-choice image quiz and its result. These quizzes are generated according to the learner's profile, location, time and the results of the past quizzes they took.
4. E-book System

The Japanese government has announced that it plans to introduce e-books in all K-12 schools by 2020. Therefore, it is a natural direction to introduce the e-book system to various kinds of learning environments. In our research, we try to link e-textbook learning with learning though real life experience (Fig. 6), to link learners’ present learning with their past learning and to link a learner’s learning with other learners’ learning using our developed system.

4.1 Architecture

E-book system is an additional system implemented to SCROLL. It runs inside SCROLL. It consists of database and EPUB files. EPUB (Electronic PUBlication) is one of the e-book formats. Fig. 7 shows its architecture. EPUB-viewer is a main function. This function shows e-Book contents to viewers, record learners’ actions as action logs, and record what learners learned from e-Book as learning logs. On EPUB-viewer, learners can take various actions, such as page turning, page jumping, bookmarking, highlighting, adding logs, taking memos, looking into the web dictionary and searching by keywords.

For instance, when they come across a new word ‘勤続 (continued service)’ while reading an article on mobile (Fig. 8 left), then click ‘Add log’ button, then it jumps to SCROLL upload interface to support their upload (Fig. 8 right).
5. Evaluation

5.1 The Target Class

The class was one of ‘international exchange subjects’ which was targeted mainly for international exchange students. Japanese students who are interested in class held in English can also join it. The target class was held 14 times once a week in a CALL (computer assisted language learning) room during the fall semester, 2016. The class language was mainly in English. The objectives of the target class were (1) to improve the skills of their target languages, which were Japanese or English and (2) to enhance cross-cultural understanding.
5.2 Method

An evaluation was conducted in one of the authors’ class at university in the western part of Japan. It consisted of 17 students (4 Japanese, 3 Germans, 2 Chinese, 2 Indonesians, 2 Taiwanese, 1 American, 1 Egyptian, 1 Hong Konger, 1 Vietnamese). All the participants were owners of mobile phones. The learning scenario is described as follows:

1. **Creating contents**
   The administrator/instructor, creates eBook contents for reading, as well as multiple-choice-quizzes using, before class.

2. **User registration**
   Students create SCROLL account. It is free.

3. **User activities**
   i) They were assigned to upload what they have learned out-of-class to the system as learning logs.
   ii) As an inside-class activity, they presented in turn what they learned out-of-class. That way their out-of-class learning was entwined with their inside-class learning through SCROLL.
   iii) As another inside-class activity, they read e-books which were implemented inside SCROLL. Since it is in SCROLL system, it powerfully support them to upload their learned words to SCROLL. E-book system

They were instructed that the number of uploaded logs reflected their grade. They were also instructed to report their out-of-class learning time and the activities they did for their target language learning to the teacher weekly. It is reported that by letting their students keep their learning reports and submit them to their teacher, it helped them get more committed to outside-class learning (Shirono, 2009). The number of SCROLL uploads and the out-of-class learning time were visualized in bar graphs weekly and shared with the whole class anonymously (Fig.9). The class was run under student-centered learning. Their presentation activities were focused in class because Dale (1946) emphasized the importance of ‘teaching others’ in terms of retention of memory (Kovalchick and Dawson, 2004). By using the term, ‘Cone of Experience’, he showed that the retention rate increased from passive learning to active learning.

![Figure 9](image_url)  
*Figure 9. The number of SCROLL uploads in bar-graph.*

5.3 Results

SCROLL was introduced in the 3rd class and it was used until the last class. The average number of SCROLL uploads was 327.5 (Table 1). As the standard deviation, 275.1 shows, the individual difference was large. The average out-of-class target language learning time was 1.3 hour per day. Their reported out-of-class learning activities included updating SCROLL, learning how to write emails (using online resources), writing reports, reviewing basic Japanese grammar material, talking to
Japanese native speaker, practicing presentation in Japanese etc. We examined if there was a correlation between the number of SCROLL uploads and the average out-of-class target language learning time. Unlike our expectation, correlation coefficient was -0.21. Therefore as far as correlation coefficient is concerned, it could hardly be said that SCROLL contributed the enhancement of out-of-class learning time in spite of the fact that most students included SCROLL activities as their out-of-class learning activity. The average out-of-class target language learning time of the most SCROLL uploader (Student #1) was as short as 0.4 hour a day and the average learning time of the least but one SCROLL uploader (Student #16) was as long as 2.5 hours a day apparently show that there was no correlation. As for low uploaders such as Student #16 and #17, they reported that they were already using other mobile app for their target language learning, they did not feel like to use SCROLL. Their passive participation influenced the whole results.

Table 1: Number of SCROLL uploads and average out-of-class target language learning time

<table>
<thead>
<tr>
<th></th>
<th>Number of SCROLL uploads</th>
<th>Average out-of-class target language learning time /day (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student #1</td>
<td>1001</td>
<td>0.4</td>
</tr>
<tr>
<td>Student #2</td>
<td>823</td>
<td>0.7</td>
</tr>
<tr>
<td>Student #3</td>
<td>743</td>
<td>1.8</td>
</tr>
<tr>
<td>Student #4</td>
<td>444</td>
<td>0.8</td>
</tr>
<tr>
<td>Student #5</td>
<td>396</td>
<td>0.9</td>
</tr>
<tr>
<td>Student #6</td>
<td>347</td>
<td>0.9</td>
</tr>
<tr>
<td>Student #7</td>
<td>332</td>
<td>1.2</td>
</tr>
<tr>
<td>Student #8</td>
<td>309</td>
<td>0.6</td>
</tr>
<tr>
<td>Student #9</td>
<td>225</td>
<td>1.4</td>
</tr>
<tr>
<td>Student #10</td>
<td>212</td>
<td>1.5</td>
</tr>
<tr>
<td>Student #11</td>
<td>209</td>
<td>5</td>
</tr>
<tr>
<td>Student #12</td>
<td>130</td>
<td>0.1</td>
</tr>
<tr>
<td>Student #13</td>
<td>121</td>
<td>1</td>
</tr>
<tr>
<td>Student #14</td>
<td>116</td>
<td>1.3</td>
</tr>
<tr>
<td>Student #15</td>
<td>113</td>
<td>1.2</td>
</tr>
<tr>
<td>Student #16</td>
<td>34</td>
<td>2.3</td>
</tr>
<tr>
<td>Student #17</td>
<td>13</td>
<td>0.7</td>
</tr>
<tr>
<td>Mean</td>
<td>327.5</td>
<td>1.3</td>
</tr>
<tr>
<td>SD</td>
<td>275.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

At the end of the phase, they were asked to answer the questionnaire as shown in Table 2. Q1 and Q2 were created based on the technology acceptance model proposed by Davis (1989). Q3 was created to examine the fun factor of our system. Q4 was created to examine the effectiveness of the quiz function of our system. Q5 was created to examine the effectiveness of the relog function. Q6 and Q7 were created for examining the user acceptence of its interface and the whole system. Q8 was created for examining the effectiveness of showing them the number of uploads weekly. Q9 was created for examining the effectiveness of SCROLL to enhance out-of-class learning.

The result of the five-point-scale survey is shown in Table 2. The highest point was given to Q8 asking about the effectiveness of showing them the number of uploads (Mean=4.2). Therefore only by showing them how many they uploaded during the week in graph encouraged them to get more involved in learning. The second highest point was given to Q1 asking about its usability (Mean=4.0), which reveals its high usability.
Table 2: The result of five-point-scale questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Was it easy for you to use SCROLL?</td>
<td>4.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Q2 Was the system helpful for you to learn Japanese words?</td>
<td>3.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Q3 Was it fun for you to use the SCROLL?</td>
<td>3.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Q4 Was the quiz function helpful for your learning?</td>
<td>3.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Q5 Did the relog function facilitate your learning?</td>
<td>3.0</td>
<td>1.3</td>
</tr>
<tr>
<td>Q6 Please rate how much you liked or disliked its interface.</td>
<td>3.5</td>
<td>1.3</td>
</tr>
<tr>
<td>Q7 Please rate how much you liked or disliked the whole system.</td>
<td>3.5</td>
<td>1.3</td>
</tr>
<tr>
<td>Q8 Knowing the number of uploads by your classmates, did it stimulate you to upload more words to the system?</td>
<td>4.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Q9 Did SCROLL contribute to increase of out-of-class learning time?</td>
<td>3.7</td>
<td>1.4</td>
</tr>
</tbody>
</table>

6. Conclusions and Future Work

The result of the evaluation showed that the whole system had a high usability. However it could hardly be said that SCROLL contributed the enhancement of out-of-class learning time. By running the student-centered active learning class, out-of-class learning could be entwined with in-class learning through SCROLL. The linking functions developed for implementation of seamless learning, such as to link e-book digital learning with real life learning by experience, to link their learned logs with related knowledge learned by other users beyond space as well as knowledge learned by themselves in the past beyond time, are expected to play an important role as a seamless learning facilitator. However, it is not in the stage where we can safely say our system facilitated Japanese language learners’ learning in the seamless learning environment. The system with enough good contents, we believe that it will attract more users and will enhance its efficiency. With the refinement of e-book function, more detail examination of effectiveness of our system will be conducted. In order to examine its effectiveness, a comparison between the pre-test and post-test will be conducted to find out if there is a statistically significant difference between them in our next evaluation experiment.

Acknowledgements

This research work was supported by Grants-in-Aid for Scientific Research No.16H06304, and No.16H03078 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References


Kukulska-Hulme, A. & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. ReCALL, 20(3), 271-289. doi:10.1017/S0958344008000335


Students’ Participative Stances and Knowledge Construction in Small Group Collaborative Learning with Mobile Instant Messaging Facilitation

Ying TANG*, Khe Foon HEW & Gaowei CHEN
Faculty of Education, The University of Hong Kong, Hong Kong
*yingtang@connect.hku.hk

Abstract: Mobile instant messaging (MIM) has become increasingly popular socially, but its educational impact on collaborative learning is still unclear. This study aims to understand how collaborative learning develops in small group discussions on a MIM platform. We collected interaction records from two groups of graduate students, who voluntarily set up private MIM groups to discuss collaborative projects. The records were analyzed with two set of codes to respectively examine levels of knowledge construction and learner participative stances. A processing mining technique was also applied to visualize how knowledge was built up on different tasks. Results suggested that MIM is probably most helpful with planning tasks and increasing interactivity with the facilitation of the pop-up notification. Informal leaders emerged in the discussions, who actively maintained group dynamics and also expressed ideas. Tasks requiring evaluation and creation might invite more higher-order knowledge co-construction.

Keywords: mobile instant message, knowledge construction, participative stance, collaborative learning

1. Introduction

Mobile devices such as smartphones have opened new possibilities for ubiquitous learning with increasing connectivity and communication. About 90% of users’ smartphone time been spent on using apps (Chaffey, 2016). Social networking, particularly mobile instant messaging (MIM), is probably the most popular category, represented by applications such as WhatsApp, Facebook Messenger and WeChat. Unlike short message service (SMS) which transmits messages through a telecommunication carrier and thus incurs a fee for the sender, MIM transmits messages through the Internet, and is therefore free to use whenever a Wi-fi connection is available. It is accessible across various platforms. Users can synchronize chat records seamlessly across devices. In addition, it supports multimodal information transmission and easy private or group discussion management. In particular, MIM is functionalized with “pop-up” notifications, with which users will know immediately when messages arrive. People can initiate a real-time communication immediately as in a synchronous talk, or they can produce asynchronous dialogues with time lag. Therefore, MIM-enabled interaction is often referred as “quasi-synchronous” (Garcia & Jacobs, 1999).

Despite its popularity, MIM remains the least explored mobile service in educational research. In particular, one area worth exploration is its possible influence on small-group collaboration, when students interact and share resources and learning experience on the MIM-enabled platform. Educators have examined the use of other technologies and their impact on collaborative learning, such as Facebook, wiki, and Twitter, but little attention has been put on MIM. This study aims to examine how students collaborated to accomplish group projects in voluntarily established MIM groups. Specifically, we intend to seek the answers in twofold: a) how knowledge was co-constructed, and b) what roles were played by students in the MIM group discussions.
2. Literature Review

Learning happens with negotiation and internalization of meaningful resources. Therefore, students are encouraged to participate in active interaction and collaboration (e.g. Laurillard, 2002), which requires them to think, plan, express, converse, negotiate and ultimately, be able to critically think and reflect (Dillenbourg, 1999). One pedagogical strategy inspired by collaborative learning in higher education is group-based assignment. Students are usually divided into small groups with normally 3 to 5 people, and asked to work on a common project. Technologies, especially mobile services, have further advanced collaborative learning with flexibility and multi-modality. In recent years, researchers have just begun to examine the use of MIM for teaching and learning purposes. We will review extant research regarding the impact of using MIM on collaborative learning.

In Fattah’s (2015) study, 15 students constructed a writing piece together based on peer feedback. The results of pre-post test revealed that students in the treatment group outperformed those in the control group in their writing skills, specifically about punctuation marks, sentence structure and idea generation. This study, however, mainly focused on terminal learning outcomes, i.e. the test results. No exploration was done on how students used MIM to collaboratively accomplish a task. Therefore, our understanding towards the dynamics of collaboration is limited. In another study, Kim, Lee, and Kim (2014) divided students into three discussion groups, using mobile IM, computer IM and discussion board, and assigned them with the same task. The results showed that even though MIM was conducive for communication, it yielded lower taskwork scores compared to the other two groups. Taskwork in this study was defined as how well students accomplished a problem-solving task in relation to elements such as novelty, importance, and relevance. However, the study merely analyzed students’ posts in terms of cognitive/metacognitive, social/interactive, and other types of interactions. The specific roles students played during the MIM discussions were not explored.

Both studies above examined the use of MIM in groups with around 15 students, which is not representative considering group work usually involves only 4-5 members in higher educational settings. Therefore, the results lack transferability into other scenarios. Another limit is these two studies adopted the use of MIM as a mandatory task with a prescribed instructor-defined procedure. By far, only one study by Miller (2016) asked four students to form a group voluntarily and create a digital scientific documentary together. Students voluntarily set up WhatsApp groups to share images and provide feedback. However, the WhatsApp interactions was not the focus of this study, and no further analysis was provided regarding the collaborative process.

To summarize, three major research and practice gaps have been identified. First, very limited number of studies examined using MIM for collaboration. Second, previous studies investigated the use of MIM in group collaboration in pre-determined research/instruction contexts. Therefore, the interaction is not initiated voluntarily by students. The groups were not assigned to the general small size of 4-5 people for group tasks either. Third, existing literature focused more on the final learning outcomes by assessment such as post-class tests. Less attention has been given on examining how students actually collaborated to complete their group task. This study tries to bridge these gaps, by examining the process of how students collaborated in voluntary MIM-enabled small groups to co-construct knowledge and complete tasks. The central research question of our study is: How is collaboration developed in self-initiated MIM-enabled small groups? Specifically, the following sub-questions have been proposed.

1. What activities were performed in self-initiated MIM-enabled small groups?
2. What participative roles emerged in self-initiated MIM-enabled small groups?
3. How was knowledge constructed in self-initiated MIM-enabled small groups?

3. Methodology

3.1 Research Context

Participants were two groups of first-year educational master students undertaking a disciplinary course in a university in Hong Kong. One required assignment was to give a group presentation on
one key adult learning strategy and design a role-play scenario to enact the chosen strategy. Students formed five small groups on their own choices to cover five strategies respectively, and each group was randomly assigned with one strategy. The assignment included introduction to the main concept, role-play to enact the strategy, discussion on the pros and cons, and proposal of implementation guidelines. All groups were given six weeks to accomplish the task.

Table 1: Codebook for knowledge construction

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-task-oriented communication (Veerman &amp; Veldhuis-Diermanse, 2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planning</td>
<td>Arranging operational details to accomplish the tasks collaboratively</td>
<td>“BTW, just would like to reconfirm that in coming few days we will collaboratively work online thru WeChat and Google drive to finish our preparation &amp; PPT, right?”</td>
</tr>
<tr>
<td>Technical</td>
<td>Asking for help or discussing technical difficulties</td>
<td>“I don't have access to the folder, I don't know why.”</td>
</tr>
<tr>
<td>Social</td>
<td>Expressions for pure social purposes, to maintain the group cohesion and inter-personal relationship.</td>
<td>“Nice. Thank you!”</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>Discussions that are irrelevant to the assignment</td>
<td>“I have a lot of reading need to be done for the other course.”</td>
</tr>
<tr>
<td>Levels of knowledge construction (task-oriented communication) (Gunawardena et al., 1997)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase I: Sharing/comparing of information</td>
<td>1) statement of observation, 2) statement of agreement, 3) corroborating examples, 4) clarification of details of statements, 5) definition of description of a problem.</td>
<td>“A community of practice is a group of people who share a concern or a passion for something they do, and learn how to do it better as they interact regularly.”</td>
</tr>
<tr>
<td>Phase II: Dissonance or inconsistency among ideas</td>
<td>1) statement of disagreement, 2) clarification of the source and level of disagreement, 3) restating position with illustration</td>
<td>“Actually, I think the role setting can be decided later, maybe we need to decide all the key concepts related to our topic at first.”</td>
</tr>
<tr>
<td>Phase III: Negotiation of meaning/co-construction of knowledge</td>
<td>1) negotiation of meaning and importance, 2) identification of overlap among conflicting concepts, 3) compromise, 4) integration</td>
<td>“I think both scratch and paper folding lack a very realistic setting. Yet, with paper folding, everyone would be more confident during presentation.”</td>
</tr>
<tr>
<td>Phase IV: Testing and modification of co-construction</td>
<td>testing the co-constructed knowledge against existing fact, understanding, experience, data and literature.</td>
<td>“Edmodo is easier after testing out. You only need to register. Google classroom requires you to set up G suit first.”</td>
</tr>
<tr>
<td>Phase V: Agreement statement/application of newly constructed knowledge</td>
<td>summarizing and applying co-constructed knowledge, meta-cognitive statement on thinking and understanding.</td>
<td>“Up until now, let me summarize what we’ve already discussed: the professor wants us to have…”</td>
</tr>
</tbody>
</table>

The topic of group A (n = 5) was “Self-directed Learning”, and for group B (n = 6), it was “Workplace Learning”. These two groups were chosen because they both contained a mixture of full-time and part-time students. The part-time students were local school teachers with busy schedules, thus it was inconvenient for them to meet face-to-face with their full-time student peers. In order to facilitate collaboration, both groups voluntarily set up their own WeChat discussion groups for communication and sharing. Students chose WeChat probably because it is the most
popular MIM app in the Asian market, especially in China, and they were using it daily for social purposes. Since students constructed the groups on their own, the instructor exerted no influence on the group interaction and collaboration.

3.2 Data Collection

Because students voluntarily set up the groups, neither the researcher nor the instructor was in their groups during the interaction and collaboration. Therefore, data collection was not carried out until the project was due. One technological affordance of WeChat is that it automatically saves interactive records unto user device or server, and makes the records easily retrievable. Therefore, we were able to collect students’ group chat history after the course ended. The chatlog recorded the sender, time and content of each message, thus so we could observe participants’ “behavior” in the computer-mediated communication usage (Mann & Stewart, 2000). As the externalization of thoughts, students’ written communication reflected their cognitive thinking process, and thus helped us understand how they collaborated with one another to accomplish the given task.

Table 2: Codebook for participative roles in small group collaboration.

<table>
<thead>
<tr>
<th>Roles</th>
<th>Definition</th>
<th>Possible behaviors</th>
<th>Actual examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Captain</td>
<td>The member puts a lot of effort to manage discussion, refer to other’s work and facilitate task accomplishment.</td>
<td>He/she usually sends messages characterized by positive tone, encouraging comments and with the aim to seek consensus.</td>
<td>“I’m sending an email to ask the instructor further on how we should select an object for training...”</td>
</tr>
<tr>
<td>Over-rider</td>
<td>A person who has strong self-oriented motivation, and stresses on his/her work and opinion.</td>
<td>He/she cares about the contributing to the groupwork by expressing ideas, yet focuses less on the group dynamics. Sometimes he/she re-stresses their points even though consensus has been reached.</td>
<td>“I know, but I don't agree with using programming for both.”</td>
</tr>
<tr>
<td>Free-rider</td>
<td>The member who aims to get benefits from groupwork (such as grades), but invests limited effort.</td>
<td>He/she asks questions and expects answers. He/she is quick to agree “orally” with little constructive information, and sometimes slow to take actions.</td>
<td>“Thanks David (pseudonym)! Will check it tonight!”</td>
</tr>
<tr>
<td>Ghost</td>
<td>The member who is strongly self-oriented, but shows low participation.</td>
<td>The participation is unrelated to the discussion, or aims to demonstrate his/her interest.</td>
<td>Little participation and contribution.</td>
</tr>
</tbody>
</table>

3.3 Data Analysis

The unit of analysis was the individual idea communicated rather than the entire piece of posting. No posting contained more than one unit of analysis. 171 units were identified in the discussion of group A, while group B produced 703 units of analysis. Two coding schemes were applied to respectively evaluate knowledge construction and participative stances. 10% of the data was coded by an independent coder for reliability check, and the inter-rater reliability reached 90%. All disagreement was solved by discussion between coders. The purpose of the first coding framework (Table 1) was to examine the levels of knowledge co-construction. The codebook was developed based on Veerman and Veldhuis-Diermanse's (2001) and Gunawardena, Lowe and Anderson's (1997) works.

The second codebook (Table 2) was based on the conceptual framework of “participative stances” in a CSCL environment (Strijbos & De Laat, 2010). We adopted this codebook to see if any roles emerged in the small group discussion and how participants interact. “Ghost” was assigned to
the participant if he/she contributed less than 10% of the communication with little constructive effort. It was possible one person demonstrated several individual stances in different situations.

4. Findings

4.1 What Activities were Performed in Self-initiated MIM-enabled Small Groups?

By calculation, we noticed that 80% of communication by group A was non-task-oriented, while only 24% of group B fell under this category. Within non-task-oriented communication, the top use was planning, such as progress management, division of labor and reminder of meetings. However, when it came to task-oriented communication, 83% communication of group A was information sharing (Phase I), while 53% of group B was on conflict statement and negotiation (Phase II to III). Phase II to Phase V communication is regarded as higher-level knowledge construction (Hew & Cheung, 2011). Therefore, group B demonstrated higher proportion of higher-level knowledge construction than group A. Figure 1 visualized the distribution of each category and showed differences between groups.

Between groups, group A mainly utilized MIM to conduct non-task-oriented activities (80%), while Group B used it for task-related purposes (76%), including sharing opinions and documents, expressing disagreement, and negotiating dissonancy. Between activities, planning was the top category in non-task-related communication for both groups, including sending reminders, confirming deadlines and setting up schedules. The most frequently demonstrated behavior in knowledge construction was sharing, including opinions, hyperlinks, and files of resources.

4.2 What Participative Roles Emerged in Self-initiated MIM-enabled Small Groups?

No participant was coded as ghost in either group based on the definition. The preliminary result showed that the most representative stance was over-rider in both groups, meaning participants actively expressed their opinions to construct knowledge, particularly more than free-riders, who invested insufficient effort into group work and tried to “ride” along with others.

However, group A had higher portion of captains and free-riders, while members in group B produced more information that characterized the members as over-riders. This result suggested that in group A, the work was more straightforwardly planned and divided with less disagreement and negotiation, while in group B, more participants were inclined to express opinions and handle the collaborative work. Further analysis of proportion of roles played in the non-task and task-related communication also corroborated this postulation. The non-task captain, who coordinated group dynamics and proposed work plans, represented higher portion in group A. Similarly, free-riders who mainly expressed agreement, also has higher representation in group A. Figure 2 was presented to better visualize the participative stances distributed in two groups.
We then examined the participative roles demonstrated by each member respectively. Both groups had one outstanding member (A1 & B1) who played a major part in each role. This person not only showed concern over the group collaboration by shouldering the duty of a captain to help with reaching consensus, but also actively made personal contribution by providing ideas (Figure 3).
4.3.1 Topic 1: Decoding the Logistical Details of the Task: What is Expected in the Presentation?

Figure 4 shows the development process of how the two groups decoded the requirements of the project, such as what the required components of the project were, how long each component would last, and in what sequence should the components be presented. To illustrate the workflow, for example, in the workflow of group A, when A2 brought up the topic, immediately A4 and A5 commented. There were three cases of disagreement, one was followed up by immediate additional supporting evidence, one was presented with no conclusive solutions, but followed by information sharing, and the other was solved by active negotiation of seeking common understanding.

Another example from group B was presented below for illustrative purpose. This excerpt was started by B4 with a question, and followed by B1 with a Phase V summary. B1 then continued summarizing what had been discussed by the members in the previous communication with three more posts, and ended this piece of excerpt by sharing another requirement of the task.

B4: So what is our focus?
B1: Focus: displaying different kinds of workplace learning...?
B1: Maybe show some issues of workplace learning as well in the role-play…
B1: What issues of workplace learning are there...........?
B1: Roles: teachers (and later on the presenters of this topic after the role-play), a mentor for interns, colleagues
B1: After the role-play, we still need to do a short presentation in the traditional way.

For group A, the discussion lasted for 8 days, with five peaks when group members were discussing at the same time. For group B, the discussion on this topic lasted for only 51 minutes. Despite the differences, the median time intervals for both groups were tagged as “instant” according to the Disco event log.

Figure 5 showed the workflow of how learners attempted to figure out the definition and meaning of the learning strategy assigned. The conversation was started by one member, and contributed by others cumulatively. Learners built on each other’s contributions mainly by sharing resources found on internet and other course materials. No conflict or negotiation was identified in either group. The major way of contribution is “Phase I: Sharing”, as shown in the excerpt below:

A3: Here are some example of SDL (self-directed learning) in primary school! I think it gives a framework for us to develop our SDL presentation
A1: Self-directed learning & andragogy [link]
A1: Don't worry, it's just the first chapter, NOT all 95 pages
For group A, the discussion sustained for just one day, while the time span of group B lasted over 3 days. However, the event log of group B showed five obvious peaks when group members gathered together online and clarified the concept of learning strategy to be presented. Although the longest interval between posts on this topic was longer (group A: 24hrs; group B: 25hrs), the majority of the interactions happened quickly online—the median interval between posting for both groups is “instant” according to the Disco event log.

Figure 5. Workflow 2: Understanding the Strategy

4.3.3 Topic 3: Designing and Evaluating the Role-play Scenario

Figure 6 showed the process of how learners designed and created a scenario to implement the learning strategy in real class teaching. The task required each group to conduct a role-play to show an actual lesson with the learning theories being implemented. Two groups demonstrated sharply different communicative patterns on creating the role-play scenarios. For group A, only two records were identified on this topic, while group B conducted an intense discussion with substantially more higher-order knowledge co-construction being invoked. For group B, the most frequently represented pattern on this discussion was: (a) one learner presented his/her own thoughts; (b) some other participants disagreed with the initial ideas/comments, or asked for further clarification; (c) the disagreement was either solved by trying to reach a quick consensus, failing which a further round of criticism and justification came into play. Despite the fact that this topic
took longer than the other topics, the median time interval between two posts is still short, ranging from instant to 3 minutes.

In the excerpt below, B4 started the communication by expressing disagreement. Rather than simply refuting, B1 summarized the task requirement details to persuade other members to accept her proposal.

B4: You have said that Scratch is very simple is it practical that a company would ask all employees to switch to something of lower standard.

B1: Flash and Powtoon are lower than scratch in some sense.

B1: And becuz scratch is newer.

B1: So let me summarize, the instructor wants us to have 1) ideally a realistic workplace issue and 2) a topic that some of the groupmates have real experience.

B1: I think both scratch and paper folding lack a very realistic setting. Yet, with paper folding, everyone would be more confident during presentation.

B4: Everyone would be confident? U mean us or our audience? I don't think group mates here are comfortable with paper folding…

5. Discussion

The conflict of schedule and limitation of physical distance are often regarded as the barriers to facilitate collaborative learning, but it seems MIM has provided possible solutions to overcoming these barriers. The cases in this study were groups set up voluntarily by students with no teacher interference. Collaborative work is nothing new in higher education, but traditionally it is only visible through the final artifact presented. By reviewing the interactive records on the MIM platform, this study hopes to reveal an otherwise neglected picture of how students collaborated to accomplish a given group task.

Based on the results of our study, both groups used MIM for planning for non-task related purposes, which required immediate decision making. The push notification of MIM has probably facilitated the planning process. Whenever a new message comes, a notification will pop up on the smartphone screen to remind the user about the arrival of new message (Barhoumi, 2015). It will remain visible as a stimulus for immediate response. This functionality has supplement the pitfall of asynchronous communication, which is scolded as it leads to lengthier time to response, and not helpful for tasks that require quick group decision making (Trentin, 2010). Besides, it may also serve as an accelerant to participation and interactivity. As nobody was found “lurking” in either group, using MIM may be helpful to build a sense of community (Wang, Fang, Han, & Chen, 2016) and encourage users to participate and interact.

As to how students performed the collaboration on the MIM platform, we examined the process of work completion, the roles of participants and knowledge construction process. Both groups had a “kernel” person (A1 & B1), who could be identified as the top figure in all participative stances. Participants got involved with no presumed roles, but these two members gradually shouldered the role of “leader” in the group communication, who would summarize previous discussion content, ask for everybody’s idea and also propose solutions to existing problems. These two students were to some degree similar to the student facilitators (Hew & Cheung, 2012), who were assigned the duty of facilitating asynchronous online discussions. Informal leaders or facilitators are considered helpful in groupwork completion. Therefore, teachers may make suggestions accordingly for student voluntary grouping, such as include someone who demonstrate such proactivity and leadership.

Besides, more higher-order knowledge construction was conducted for more demanding tasks, such as “to evaluate” or “to create”. As Schellens and Valcke (2005) discussed, the nature of task matters when teachers try to encourage students to have higher phases of knowledge construction. This information might be inspirational for task design. Teachers may take the task nature into consideration to help students benefit more from higher-order thinking.
6. Limitation

There are several limitations in this study. First, the two groups demonstrated major differences in regard to data size and data features, yet we were not able to conduct follow-up interviews with the participants to seek explanations to the disparity, for some objective reasons. Second, the current study only examined learner collaboration using interactive records. Future study may triangulate the data, such as students’ self-reported data and group work artifacts and scores. Third, we only collected data from two groups from the same post-graduate class. The transferability is therefore limited. Fourth, the current study did not have comparison groups to confirm the effect of using MIM for collaborative learning. Future study may compare face-to-face learning, MIM, and online forum, and comprehensively understand the impact of different communicative mode on students’ interaction and learning.

References


Learning Behavioral Pattern Analysis based on Students’ Logs in Reading Digital Books

Chengjiu YIN*, Noriko UOSAKIb, Hui-Chun CHUC, Gwo-Jen HWANGd, Gi-Zen Liu e, Jau-Jian HWANGf, Itsuo HATONOG, Etsuko KUMAMOTOC & Yoshiyuki TABATAf

a Information Science and Technology Center, Kobe University, Japan
b Center for International Education and Exchange, Osaka University, Japan
c Computer Science & Information Management, Soochow University, Taiwan
d College of Liberal Arts and Social Sciences, National Taiwan University of Science and Technology, Taiwan
e Department of Foreign Languages & Literature, National Cheng Kung University, Taiwan
f Research Institute for Information Technology, Kyushu University, Japan

*yin@lion.kobe-u.ac.jp

Abstract: In this paper, we presented a study developing a digital textbook system, which could be used anywhere, anytime. An experiment was conducted using the developed system to collect students’ learning logs for analyzing their behavioral patterns on an Educational Technology course for graduate students. In the experiment, we assigned the students to read an academic English journal article. The lag-sequential analysis method was employed to analyze and infer their behavioral patterns. Several interesting behavioral patterns were found from the analysis results. The findings are helpful to the improvement of the digital textbook system; moreover, some behavioral patterns could provide helpful references for teachers to improve teaching materials in the future.

Keywords: Learning behaviors, lag-sequential analysis, learning log, digital textbooks

1. Introduction

Nowadays, smartphones have become very popular devices for communications and learning because the functions of these devices have been significantly improved in recent years and the physical size and weight have been better designed for increasing portability (Yin et al., 2016). By using mobile devices, digital textbooks can be conveniently accessed through Internet anywhere and anytime. With the development of e-publishing technologies and standards increasingly, more and more traditional textbooks have been replaced by digital ones (Rainie, et al., 2012, Yin et al., 2014).

In the past decade, various studies have been conducted to investigate the effectiveness of learning with digital textbooks. For example, Shepperd et al. (2008) compared efficacy between digital textbooks and traditional textbooks, and they indicated that students rated the usability of digital textbooks positively. Rockinson-Szapkiw et al. (2013) compared the learning effectiveness between digital textbooks and traditional textbooks. They found that digital textbooks are as effective for learning as traditional textbooks. Therefore, many researchers concur that digital textbooks have become a potentially effective pedagogic tool for supporting teaching, learning and scholarship (Hezroni, 2004; Reinking, 1997; Snyder, 2002).

In the meantime, learning analytics has become an important issue in education. Learning analytics plays an important role in providing helpful suggestions to policy makers, instructors or learners by analyzing learning logs or educational data (Baker & Inventado, 2014; Hwang et. al.,2017). The objective of learning analytics is to provide helpful information to optimize or improve learning designs, learning outcomes and learning environments based on the analysis results (Greller & Drachsler, 2012).

In this paper, a digital textbook system is developed to collect data in the classes. The system is named Digital textbook for Improving Teaching and Learning (DITeL). The DITeL system can be used...
not only on personal computer, but also on smart phone. That is to say, this digital system can be used anywhere, anytime. Teachers and students can use DIoTeL system and read the digital textbook by using mobile devices such as iPad, iPhone, and Android. And their learning logs were collected for analyzing their learning behaviors to improve DIoTeL system.

In order to analyze learning behaviors, we designed an experiment using DIoTeL system to collect students’ learning logs. The experiment was carried on Educational Technology course for graduate students. In the experiment, we assigned them to read an academic English journal article. After learning activities, we applied lag-sequential analysis to analyze and infer their behavior patterns. We found some behavioral patterns which may help to improve digital textbook system. We also found some behavioral patterns which may help teachers to improve their teaching material.

2. Literature Review

2.1 Previous Studies of Data Collection

Collecting data is the first step in learning analysis (Yin et al., 2013a; 2013b). In May 2015, we thus performed a review of previous research to survey the categories that can be classified in terms of data collection (Yin et al. 2016).

Based on the data source, previous studies on data collection could be classified into three categories: Questionnaire-based Data Collection (QDC), Manual Data Collection (MDC), and Automatic Data Collection (ADC) (Yin et al., 2014; Ren et al., 2017).

- **QDC.** In this category, data are collected by using a predesigned questionnaire. Ho et al. (2013) used a questionnaire to investigate the teacher behavior on adopting mobile phone messages as a parent–teacher communication medium.

- **MDC.** In this category, a manual data collection system is opened to users, who can employ the system and consciously provide data about their learning behaviors. For example, Chiang et al. (2014) provided an augmented reality (AR) system to guide students in knowledge sharing in inquiry learning activities. In this approach, students capture images from an authentic environment and share these with others.

- **ADC.** In this category, learning behaviors log data are automatically recorded while reading e-documents. For example, Yin et al. (2015) analyzed learning behavior and identify students' learning style using student's digital textbooks reading logs data, which were recorded automatically. By using same digital textbooks logs data, Shimada et al. (2017) summarized lecture slides to enhance preview efficiency and improve students’ understanding of the content, Mouri and Yin (2017) find some patterns for improving learning materials.

For categories QDC and MDC, the data are consciously collected. Therefore, data are affected by users’ own subjective factors. For category ADC, the data is objectively collected, thereby removing the subjective factors that affect data authenticity. The present work falls under category ADC.

2.2 Behavioral Sequential Analysis

Behavioral sequential analysis is a statistical analysis method. Through a series of sequential analysis matrix calculations to determine behavioral transitions (Bakeman & Gottman, 1997; Hou, 2012). There are many researches using a series of progressive sequential analyses to analyze learning behavioral patterns and they point out the benefit of using progressive sequential analyses (Hou, 2012; Hsieh et al., 2016; Hwang et al., 2017). Hou (2012) indicated that using a visualized behavior–transition diagram to explore learners’ complex behaviors can help to develop a more effective instructional mechanism for game-based system. Therefore, we employed this method to improve digital textbook system.
3. Digital Textbook System

By using e-pub format, a web-based digital textbook system was developed and used for this research (Fig. 1, Fig. 2). Fig. 1 is an interface for students, and Fig. 2 is an interface for teachers. By using this online digital textbook reading system, we can collect data like “turning to next/previous page”, “memo”, “zoom in/out”, “adding marker”. All of these actions are stored to the database. These data were used to analyze learning behaviors.

**Turning to next/previous page.** Students can read the teaching contents again and again, they go to the next page by clicking “Next” button, and backtrack to the previous page by clicking “Prev” button.

**Memo.** While a user want to write some memo in the learning content, he will click “Memo” button, and a textbox will be shown. After he finished writing memo, the action name will be saved as “Memo”.

**Zoom in/out.** The zoom in/out function can help students read the contents more clearly.

**Adding marker.** While a user want to highlight some text in the learning content, s/he will click “abc highlight” or “Under line” button, and the action name will be saved as “Highlight” or “Underline”.

Teacher can register each student's name and student number into the system. Before the students login in the system, the digital textbook and other relevant materials have been uploaded to the system by the teacher.

Each student will have his/her own account to enter the system, so that he/she has a separate record of this course to learn.

![Fig. 1. Student interface of DITeL](image-url)
4. Experimental Design

To analyze students’ behavioral patterns in learning with digital textbooks, an experiment was designed using DITeL system to collect students’ learning logs. The experiment was carried on an Educational Technology course for graduate students. The aim of the study was to explore the learning behaviors of students reading academic papers. The progressive sequential analysis was used to infer the learning behaviors of students when they were reading the academic papers.

4.1 Participants

A total of 21 graduate students participated in this study. The participants were asked to read an academic paper via the digital textbook system. The age of the participants was 23 on average. The experiment was carried about 1.5 hours.

To protect the participants, the experiment was conducted following the ethics criteria suggested by an authorized ethics committee in Japan. That is, the participants were protected by hiding their personal information during the research process; moreover, they knew that their participation was voluntary and that they could withdraw from the study at any time. As the result, we could use 17 participants to do learning analysis.

4.2 Coding Scheme

To do a progressive sequential analysis, a coding process is usually required. However, this study only analyzes the learning behaviors which were recorded automatically, therefore, no coding process needed to be carried out in this system. That is, the coding is based on their operating behaviors in the digital textbook system.
5. **Analysis of the Learning Behavioral Patterns**

Totally 1,370 learning behaviors were recorded.

5.1 *Analysis of the Frequency of Behavioral Patterns*

Table 1 shows the frequency and percentage of the individual coded behaviors of the students. It was found that "go to next page" (NEXT) and "go to previous page" (PREV) were the most frequent behaviors. The percentage of "go to next page" was 39%, and "go to previous page" was 28%. These were most likely behaviors, as students needed to flip to read the textbook.

We also found that marker functions, "make underline" (UNDERLINE), and "make highlight" (HIGHLIGHT), were also used frequently. However, the percentage of "make underline" (6%) is less than, "make highlight" (22%). Although both are all marker functions, the students liked to use highlight better than underline.

**Table 1: The frequency and percentage of coded behaviors of student.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency</th>
<th>Percentage(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREV (PR)</td>
<td>390</td>
<td>28</td>
</tr>
<tr>
<td>NEXT (NX)</td>
<td>537</td>
<td>39</td>
</tr>
<tr>
<td>UNDERLINE (UL)</td>
<td>79</td>
<td>6</td>
</tr>
<tr>
<td>DEL UNDERLINE(DU)</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>MEMO(MO)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>HIGHLIGHT(HL)</td>
<td>301</td>
<td>22</td>
</tr>
<tr>
<td>DEL HIGHLIGHT(DH)</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>BOOKMARKER(BM)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>DEL BOOKMARKER(DB)</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

5.2 *Analysis of the Learning Behavioral Patterns*

To probe the behavioral patterns of the students in reading digital textbooks, a series of progressive sequential analyses was conducted to explore the behavioral patterns. As shown in Table 2, the rows represent the starting behaviors, and the columns represent the subsequent behaviors. A Z-value greater than 1.96 means that a behavior-sequence reaches the level of significance (p < 0.05) (Bakeman & Gottman, 1997; Hou, 2012).

**Table 2: Sequential analyses table (n = 17)**

<table>
<thead>
<tr>
<th>Z-value</th>
<th>PR</th>
<th>NX</th>
<th>UL</th>
<th>DU</th>
<th>MO</th>
<th>HL</th>
<th>DH</th>
<th>BM</th>
<th>DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>11.67</td>
<td>-2.70</td>
<td>-3.98</td>
<td>-1.31</td>
<td>-0.83</td>
<td>-5.15</td>
<td>-3.44</td>
<td>-1.79</td>
<td>-1.09</td>
</tr>
<tr>
<td>NX</td>
<td>-1.33</td>
<td>9.95</td>
<td>-4.98</td>
<td>-3.13</td>
<td>0.20</td>
<td>-6.19</td>
<td>-1.29</td>
<td>-0.10</td>
<td>-0.21</td>
</tr>
<tr>
<td>UL</td>
<td>-4.20</td>
<td>-5.13</td>
<td>15.61</td>
<td>2.37</td>
<td>-0.66</td>
<td>1.37</td>
<td>-0.55</td>
<td>-0.70</td>
<td>2.05</td>
</tr>
<tr>
<td>DU</td>
<td>-1.89</td>
<td>-2.05</td>
<td>3.45</td>
<td>7.01</td>
<td>-0.28</td>
<td>0.44</td>
<td>1.21</td>
<td>-0.30</td>
<td>-0.18</td>
</tr>
<tr>
<td>MO</td>
<td>-1.01</td>
<td>0.64</td>
<td>-0.71</td>
<td>-0.30</td>
<td>9.66</td>
<td>-0.65</td>
<td>-0.42</td>
<td>-0.22</td>
<td>-0.13</td>
</tr>
<tr>
<td>HL</td>
<td>-6.60</td>
<td>-4.07</td>
<td>0.48</td>
<td>1.08</td>
<td>-0.49</td>
<td>10.06</td>
<td>4.38</td>
<td>0.21</td>
<td>-0.92</td>
</tr>
<tr>
<td>DH</td>
<td>-3.03</td>
<td>-2.82</td>
<td>-0.56</td>
<td>1.21</td>
<td>-0.39</td>
<td>6.18</td>
<td>0.49</td>
<td>-0.42</td>
<td>3.73</td>
</tr>
<tr>
<td>BM</td>
<td>-1.79</td>
<td>-1.54</td>
<td>2.31</td>
<td>3.08</td>
<td>-0.20</td>
<td>0.21</td>
<td>4.47</td>
<td>-0.22</td>
<td>-0.13</td>
</tr>
<tr>
<td>DB</td>
<td>-1.10</td>
<td>-1.39</td>
<td>-0.43</td>
<td>-0.18</td>
<td>-0.13</td>
<td>-0.92</td>
<td>-0.26</td>
<td>22.44</td>
<td>-0.08</td>
</tr>
</tbody>
</table>
Table 2 indicates there are 16 significant sequences that occurred during the reading digital textbooks. Based on the significant sequences, a diagram of behavioral-transition was prepared. Fig. 3 shows the behavioral transition diagrams of the students. All of the sequences in the diagrams are statically significant. The values above each line represent the z-score for the sequence, while the direction of the line represents the direction of the behavioral transition.

As shown on Fig. 3, it was beyond our expectation that PREV, NEXT and MEMO behaviors have no sequential correlations between other learning behaviors. However, they have sequential correlations with themselves (PREV → PREV; NEXT → NEXT; MEMO → MEMO). We also found some other learning behavioral patterns: LBP1-6 as shown in Table 3.

Table 3: Learning behavioral patterns.

<table>
<thead>
<tr>
<th>No</th>
<th>Learning Behavioral Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP1</td>
<td>“HIGHLIGHT” has sequential correlations with itself and “DEL HIGHLIGHT”</td>
</tr>
<tr>
<td>LBP2</td>
<td>“BOOKMARK” has sequential correlations with “DEL BOOKMARK”</td>
</tr>
<tr>
<td>LBP3</td>
<td>“BOOKMARK” has sequential correlations and “DEL UNDERLINE”</td>
</tr>
<tr>
<td>LBP4</td>
<td>“BOOKMARK” has sequential correlations with “DEL HIGHLIGHT”</td>
</tr>
<tr>
<td>LBP5</td>
<td>“DEL HIGHLIGHT” has sequential correlations with “DEL BOOKMARK”</td>
</tr>
<tr>
<td>LBP6</td>
<td>“DEL UNDERLINE” has sequential correlations with itself and “DEL BOOKMARK”</td>
</tr>
</tbody>
</table>

We have shown the students the learning behavioral patterns (Table3) and carried out an interview to ask them why they took such actions.

5.2.1 LBP1

It was found that “After adding HIGHLIGHT, the students deleted the HIGHLIGHT, or after deleting HIGHLIGHT, they added HIGHLIGHT again”; Some of the students who had these learning behavioral patterns, stated their perceptions as follows:
1. I highlighted it because I thought it was the main idea of the paragraph, but I realized it was wrong, so I deleted it.
2. Because I made a mistake in highlight position.
3. It was a mistake of my operation.
4. I highlighted on some words, after that, I found more meaningful words.
5. Because I thought it was an important place, after I read the rest of the paper, I found it was not important.
6. Because I made a mistake in the range of the highlight.

From the interview, it was found that students often changed the important keywords when they were reading the textbook, and that it was difficult for them to identify which words were important. It suggests that it would be appropriate to mark the important places on the teaching materials before students read the contents.

5.2.2 LBP2

It was also found that, after adding BOOKMARK, the students would delete the BOOKMARK; after deleting BOOKMARK, they would add BOOKMARK again. Some of the students who had this learning behavioral pattern, stated their perceptions as follows:

1. I thought it was an important page, but after I read the rest of the paper, I found it was not important, and added bookmark on another page.
2. I examined the importance of the pages again and removed those of less importance.
3. When I had some other things to do, which means I have to read the article later, I will add a new bookmark so that I can continue my work later.

From the interview, it was also found that students often changed the important page while they were reading the textbook, they were often confused about which pages were important. It is suggested that it would be appropriate to mark the important pages on the teaching materials before students read the contents.

5.2.3 LBP3, LBP4

Another finding was that, after “add bookmarker”, the students often “delete highlight” or “delete underline”. This phenomenon may suggest that there are sequential correlations between “BOOKMARKER” and marker functions (UNDERLINE, HIGHLIGHT). Some of the students who had this learning behavioral pattern stated their opinions as follows:

1. I can use highlight instead of using underlines, without any specific reasons. And there is also the bookmark function, so maybe it is not necessary to use underlines.
2. Because I thought it was not necessary to underline/highlight here. It is enough to write memo here.
3. When I found more important parts than the parts I underlined/highlighted, I added the bookmark there and deleted the underline/highlight.
4. If I add the bookmark, then it is not necessary to use underline.
5. It is not easy to use underline and highlight functions.

From the interview, we found that some students preferred bookmark to highlights/underlines. Some students preferred memo to mark. It suggests that the students preferred to use bookmark and memo functions to mark functions. And it seems that the underline and highlight functions of our system are not user-friendly. The improvement of these functions is among our future works.

5.2.4 LBP5, LPB6
It was also observed that, after “deleting highlight/underline”, the students often “delete bookmark”. Some of the students who had this learning behavioral pattern share the following opinions:

1. When I completed the reading of the paper, I felt that I understood all of them.
2. I thought the part which had been highlighted was not important anymore, so I deleted the highlight or the bookmark.
3. When I completed the reading of the paper, I felt the place that I marked was not so important.
4. When I had questions on the contents I marked it, after when I found the answer, I deleted them all.

From the interview, we found that sometimes mark functions were used temporarily, such as if they had questions on some contents, then they added marks on these contents. After they found the answer, they deleted them. Therefore, if we add links between the related contents beforehand, students can read the textbook in a more efficient way.

6. Conclusion

With the development of e-publishing technologies and standards increasingly, more and more traditional textbooks have been replaced by digital textbook. In this paper, we developed a digital textbook reading system, which could be used anywhere, anytime. Teachers can upload their teaching materials themselves and students can use the system to read textbooks. At the same time, students’ reading logs were collected to analyze their learning behaviors. In order to probe the behavioral patterns of the students in reading digital textbooks, a series of progressive sequential analyses was conducted to explore the behavioral patterns.

We found some behavioral patterns which may help digital textbook system developers and instructional designers reach an in-depth understanding of the actual operations and behavioral patterns of learners. It also enables them to use a visualized behavior–transition diagram to explore learners’ complex behaviors and develop a more effective instructional mechanism for digital textbook systems in the future.

Acknowledgements

Part of this research work was supported by the Grant-in-Aid for Scientific Research No.16H03078 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.

References


A Case Study of Evaluation of Learners’ Acceptance of AR_H₂O₂ System

Tao Wang, Shan Jia, Jinrui Dai, Manli Lu, Xiaoru Xue, Su Cai & Feng-Kuang Chiang
School of Education Technology, Faculty of Education, China
Joint Laboratory for Mobile Learning, Ministry of Education-China Mobile Communications Corporation, China
Beijing Advanced Innovation Center for Future Education, China
caisu@bnu.edu.cn

Abstract: The AR_H₂O₂ system for simulating decomposition of hydrogen peroxide with augmented reality is presented. Based on Technology Acceptance Model (TAM) enhanced with perceived enjoyment, we designed a questionnaire in order to evaluate the attitude of learners towards AR-based experiment. For empirical study, 50 students in Grade Two at Junior Middle School were involved and completed a survey questionnaire. The result of our study shows that (1) perceived enjoyment and attitude towards using have a comparable effect on the intention to use; (2) perceived enjoyment plays a dominant role in determining the attitude towards using this system; and (3) both perceived usefulness and perceived enjoyment are depended on perceived of using this system.

Keywords: augmented reality, evaluation of CAL system, Technology Acceptance Model

1. Introduction

With the development of science and technology, information technology is playing a significant role in all aspects of our life. In the contemporary days, education has taken a new turn under this influence as E-learning has driven a revolution for both IT and education. IT revolution process has met many challenges as the effects of these new technologies need to be checked.

The majority of students in middle school have weak ability to think in abstractions. This common confusion is usually settled in the laboratories. However, the experimenting rooms as well as the materials for the trials are in a shortage in most cases. When it comes to enabling learners to think in a rational way of a completely new disciplines like chemistry and physics, teachers and instructors in junior high school always found a sharp difference among students’ learning. As a result, the efficiency of the class is low and some students may fail to understand what they need to know.

Augmented reality (AR) is a technology to simulate a natural situation with realistic visual, hearing and touching effects with the support of computers and other portable devices. AR realizes direct natural interactions between the users and the environment and enables users to immerse into the environment through a variety of sensing equipment. In consideration of the situation in middle schools, there comes the need to explore whether bringing AR into classrooms can help solving the problem of shortage in experimenting resources and difference in experimenters’ individualized pace.

With the portability of AR devices, students are allowed to learn the points anywhere following their own learning paces. This convenience plays an important role in places where the laboratories are not available and deals with the challenge brought by individualized learning. AR, as a natural interactive way, can help students better involved in the experiment and corresponding knowledge as well as deepen their understanding of the experimental operations. For example, the knowledge of particles of atoms and molecules, ions and atoms such as micro structure has been regarded as an aporia in chemistry. As students found it difficult to visualize the abstract micro particle image, it becomes demanding and time-consuming for students to learn this part. The introduction of AR into classroom helps the visualization of abstract problem (Cai, Wang, & Chiang, 2014).
2. Literature Review

The application of AR to middle school teaching has become a hot issue. In secondary school textbooks, some experiments cannot be carried out because of slow reaction rates, lack of chemicals and instrumental resources. The risk of poisonous gases and explosion has to be taken into consideration when hazardous chemicals are in use. AR can solve these problems by not only describing the experimental phenomenon, but also enhancing students' interaction with experimenting devices and their partners. In this way, it is possible to improve students' self-learning and cooperation skills, provoke their thoughts and deepen their understanding during AR experiments.

The introduction of AR is helpful for students to learn disciplinary knowledge. Some researchers used AR to simulate the mechanical movement. The results show that the technology can assist students in learning kinematics (Baritz, Cotoros, & Moraru, 2007). Cai et al. (Cai et al., 2016) combined the AR and Kinect equipment to design the motion-sensing instructional software to help students learn physics in high school. The results show that AR can improve the efficiency and effectiveness of students' learning.

At present, it is found that AR would affect students' motivation and interest for learning. Chang et al. (Chang, Wu, & Hsu, 2013) carried out their study to support students to learn in hot spots of society like nuclear energy. The study found that students' perceptions of AR activity have a vital influence on the shift in attitudes toward nuclear and that AR may affect the learner's emotional attitude towards real-world problems. Di Serio et al. (Serio, Ibáñez, & Kloos, 2013) developed Instructional Materials Motivation Survey (IMMS) based on the ARCS motivation model. The results indicate that AR has a positive impact on the motivation of middle school students.

Previous studies have explored a lot on the influence of AR system on students' academic performance, learning interest and motivation while few of them focused on students' attitude. It is obvious that the students' attitude towards AR have an important impact on the AR system's use, maintenance, promotion and other aspects. There are some relevant researches on students' acceptance of AR. Wojciechowski and Cellary (2013) used ARIES system to set up the scene of a chemistry experiment. The questionnaire to evaluate the experimenters' attitude towards learning in ARIES AR environments has been designed as well. Technology Acceptance Model (TAM) expanded dimensions of measurement to perceived enjoyment. By constructing friendly interfaces, it is found that perceived usefulness and enjoyment had a comparable effect on the learners' attitude toward using AR environments. Perceived enjoyment played a dominant role in experimenters' determination on whether to use AR. Ibanez et al. (Ibanez, Serio, Villaran, & Delgadokloos, 2016) investigated the attitude of learners towards an AR activity that is designed to help engineering students solve electromagnetic problems. Based on TAM model, the results of the evaluation show that the intention to use the system is dependent on perceived enjoyment, rather than on perceived usefulness of the learning tools. Students' technology acceptance towards the system is also an important measurement. However, owing to the lack of study in this field, there does not exist such a systematic model measuring the influence of students' acceptance towards the new technology.

3. System Design

In this section, a tool for building AR-based learning environments, called AR_H2O2, is presented. AR_H2O2 is developed to simulate decomposition of hydrogen peroxide with AR technology by supplying learners with vivid real-time demos. Therefore, AR_H2O2 can be used as an alternative tool to help learners in rural areas where cannot meet the experimental to do experiments.

3.1 AR System Introduction

AR_H2O2 contains AR software and four markers. The software contains three processes of hydrogen peroxide decomposition, and the effects of reaction temperature, concentration and catalyzer on decomposition rates are investigated respectively.

The markers are used to interact with this software. Each set contains four markers printed with particular textures, which are selectively applicable to different applications. When the software runs,
learners can manipulate markers to investigate the effects of different factors on the reaction rates referring the documentation and further generalize concepts and conclusions.

3.2 Design and Development of AR_ H2O2

The development of AR_ H2O2 can be divided into three phases: capture real scene, track and compare the marker, and compositing rendering, as shown in Figure 1. The models built-in the software is built in 3DS Max platform, and software is programmed by using Unity3D editor and scripting using C#. Besides, we adjust the co-ordinate system and the interactive mode between users and the models. AR software development kit (SDK) called Vuforia is implemented to make sure AR_ H2O2 can render the real scene and virtual models to create a real-time mixed reality environment. As a convenience to learners, we install virtual button in specific areas, which could help learners to manipulate the reaction factors (including the valve of temperature, concentration and catalyzer), while the valve of reaction factors is shown in label built-in the soft, as shown in Figure 2. Figure 2 (a)(b)(c) shows process of hydrogen peroxide decomposition which is influenced by temperature, concentration and catalyzer respectively.

4. Description of Experiment

4.1 Research Model

The aim of the experiment is to evaluate the learners’ attitude towards the AR devices. Therefore, this study adopts the Technology Acceptance Model to explore the determinants of learners' acceptance toward AR devices. Technology acceptance models are widely used in technology acceptance studies.

Perceived usefulness is defined as “the degree to which a person believes that using a particular system would enhance his or her job performance”. In our research, it represents the user believes that the AR system would yield positive benefits for chemistry learning. Perceived ease of use refers to “the degree to which a person believes that using a particular system would be free of effort” (Davis, 1989). In our study, it means the operation of using this system for chemical learning is simple and clear.

Davis et al. proposed an improved version of the Technology Acceptance Model, suggesting perceived enjoyment as an intrinsic motivation indicator of attitudes. Perceived enjoyment is defined as “the extent to which the activity of using the computer is perceived to be enjoyable in its own right, apart from any performance consequences that may be anticipated” (Davis, Bagozzi, & Warshaw, 1992). Many studies have confirmed that perceived enjoyment has a significant effect on attitudes, so perceived enjoyment should be considered as an indicator in the TAM model (Dickinger, Arami, & Meyer, 2008; Sun & Zhang, 2005; Teo & Noyes, 2011).
In this study, we selected the Davis’ enhanced version of TAM model with perceived enjoyment. The indicators that impact students’ attitude of the AR system are presented in Figure 3. According to the model, perceived usefulness and perceived enjoyment directly influence attitude toward using and intention to use the system. Furthermore, perceived ease of use may directly affect perceived usefulness and perceived enjoyment, and the attitude toward using. At the same time, the intention to use is directly affected by perceived usefulness, perceived enjoyment and the attitude towards using.

The following research hypotheses were formulated on the basis of the research model:

- H1. Perceived ease of use (PEU) will positively affect perceived usefulness (PU).
- H2. Perceived ease of use (PEU) will positively affect perceived enjoyment (PE).
- H3. Perceived ease of use (PEU) will positively affect attitude toward using (ATU).
- H4. Perceived enjoyment (PE) will positively affect attitude toward using (ATU).
- H5. Perceived usefulness (PU) will positively affect attitude toward using (ATU).
- H6. Perceived usefulness (PU) will positively affect intention to use (ITU).
- H7. Perceived enjoyment (PE) will positively affect intention to use (ITU).
4.2 Participants

50 students of 13-15 years old from a middle school were chosen in this study. The chemistry curriculum for the third year of junior high school includes factors that influence the rate of chemical reactions. Therefore, the experiments chosen in this study involve the factors that influence the rate of chemical reaction, which is consistent with the student's current curriculum.

4.3 Procedure

At the beginning of the experiment, the AR technology was introduced as a course introduction. After that, students filled out a 15-minute test paper about the factors influencing the reaction rate of the decomposition of hydrogen peroxide. Then, as shown in Figure 4, three students form a group using the AR system to learn chemical knowledge as well as doing the relevant virtual experiments. All 50 participants completed the experiment successfully. After completing the experiment, the students completed the same chemistry test as before. After the test, the participants were asked to fill out an anonymous questionnaire with statements about working with the AR system and the attitude toward using such a system in the learning process in the future.

According to the research model in Figure 3, we designed and developed the research questionnaire. The participants needed to fill in personal information and 15 questions were divided
into five groups representing the indicators of the research model. The questionnaire referred to the relevant questionnaire of the TAM model in the previous study and was adjusted according to the AR environment. Each statement in the questionnaire was measured according to a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with the exception of one reversed item for attitude toward using, which was measured in a five-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree).

5. Data Analysis and Results

5.1 Procedure

The statements of the questionnaire and the descriptive statistics for each statement are presented in Table 1. All mean values are within a range of 2.78 and 4.48. The standard deviation ranges from 0.526 to 1.577.

<table>
<thead>
<tr>
<th>Questionnaire statements</th>
<th>M</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived usefulness (PU)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The use of such a system improves learning in the classroom.</td>
<td>4.48</td>
<td>0.847</td>
</tr>
<tr>
<td>Using the system during lessons would facilitate understanding of certain concepts.</td>
<td>4.48</td>
<td>0.784</td>
</tr>
<tr>
<td>I believe that the system is helpful when learning.</td>
<td>4.53</td>
<td>0.816</td>
</tr>
<tr>
<td><strong>Perceived ease of use (PEU)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think the system is easy to use.</td>
<td>4.53</td>
<td>0.751</td>
</tr>
<tr>
<td>Learning to use the system is not a problem.</td>
<td>4.70</td>
<td>0.648</td>
</tr>
<tr>
<td>Operation with the system is clear and understandable.</td>
<td>4.68</td>
<td>0.526</td>
</tr>
<tr>
<td><strong>Perceived enjoyment (PE)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think the system allows learning by playing.</td>
<td>4.45</td>
<td>1.037</td>
</tr>
<tr>
<td>I enjoyed using the system.</td>
<td>4.45</td>
<td>0.959</td>
</tr>
<tr>
<td>Learning with such a system is entertainment.</td>
<td>4.70</td>
<td>0.758</td>
</tr>
<tr>
<td><strong>Attitude toward using (ATU)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The use of such a system makes learning more interesting.</td>
<td>4.60</td>
<td>0.744</td>
</tr>
<tr>
<td>Learning through the system was boring (reversed item).</td>
<td>2.78</td>
<td>1.577</td>
</tr>
<tr>
<td>I believe that using such a system in the classroom is a good idea.</td>
<td>4.60</td>
<td>0.744</td>
</tr>
<tr>
<td><strong>Intention to use (ITU)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would like to use the system in the future if I had the opportunity.</td>
<td>4.45</td>
<td>0.904</td>
</tr>
<tr>
<td>Using such a system would allow me to perform chemical experiments on my own.</td>
<td>4.50</td>
<td>0.784</td>
</tr>
<tr>
<td>I would like to use the system to learn chemistry and other subjects.</td>
<td>4.52</td>
<td>0.751</td>
</tr>
</tbody>
</table>

To measure the internal consistency of statements, a coefficient Cronbach alpha was calculated for the statements belonging to each construct specified in the research model. To consider the internal reliability of statements concerning the same construct as satisfactory Cronbach alpha should be greater than 0.7. The obtained Cronbach alpha values for each construct except ATU are at a satisfactory level, as shown in Table 2. In the case of ATU, the value is slightly lower, which may indicate minor differences between the statements formulated regarding attitude toward using. This discrepancy could be influenced by the fact that one of the three statements was a reversed item phrased in the opposite semantic direction from the other statements. Negative statements used together with positive statements can decrease the degree of internal consistency, because the negative items may not be considered the exact opposite of the positive ones.
Table 2: The Cronbach Alpha Values

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cronbach alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness (PU)</td>
<td>0.858</td>
</tr>
<tr>
<td>Perceived ease of use (PEU)</td>
<td>0.788</td>
</tr>
<tr>
<td>Perceived enjoyment (PE)</td>
<td>0.845</td>
</tr>
<tr>
<td>Attitude toward using (ATU)</td>
<td>0.351</td>
</tr>
<tr>
<td>Intention to use (ITU)</td>
<td>0.892</td>
</tr>
</tbody>
</table>

5.2 System Evaluation

In order to validate hypotheses H1 and H2, we used regression analysis to examine the relationships between pairs of the appropriate constructs defined in research model. The results of the regression analysis are presented in Table 3.

Table 3: The results of regression analysis

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Coef</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness (PU)</td>
<td>Perceived ease of use (PEU)</td>
<td>0.906</td>
<td>0.454</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Perceived enjoyment (PE)</td>
<td>Perceived ease of use (PEU)</td>
<td>0.998</td>
<td>0.435</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Following the empirical study, for the hypotheses H1 and H2 we rejected the null hypothesis denoting the lack of dependence, since the p-value was less than the assumed significance level of 0.05. Perceived usefulness was dependent to perceived ease of use ($R^2 = 0.454$) and perceived enjoyment also depends on perceived ease of use ($R^2 = 0.435$). Therefore, the hypotheses H1 and H2 were supported based on the regression values.

For the purpose of studying factors that influence attitude toward the using and intention to use, we used stepwise multiple regression analysis to obtain the best models. The results of the stepwise multiple regression analysis are shown in Table 4.

Table 4: The result of the stepwise multiple regressions analysis

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Coef</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention to use (ITU)</td>
<td>Perceived enjoyment (PE)</td>
<td>0.536</td>
<td>0.625</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Attitude toward using (ATU)</td>
<td>Perceived enjoyment (PE)</td>
<td>0.294</td>
<td></td>
<td>&lt;0.03</td>
</tr>
<tr>
<td>Attitude toward using (ATU)</td>
<td>Perceived enjoyment (PE)</td>
<td>0.525</td>
<td>0.329</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The results of the stepwise multiple regression analysis show that intention to use depends on perceived enjoyment and attitude towards using ($R^2 = 0.625$), supporting the hypotheses H7 and H8. Perceived ease of use is excluded by stepwise multiple regression analysis because of the high p-value ($p = 0.88$). This suggested that the hypothesis H6 was not supported. As a result, user’s perceived enjoyment and attitude towards using positively affect their intention to use. Besides, based on the step-wise multiple regression analysis, attitude towards using was depended on perceived enjoyment ($R^2 = 0.329$), indicating the hypothesis H4 is supported. From the model, attitude towards using is expected to increase by 0.525 when perceived enjoyment increases by one.

6. Conclusion

It is concluded that perceived ease of use has a significant impact on perceived usefulness and perceived enjoyment. The attitude towards using is directly related to perceived enjoyment, and the intention to use is influenced by perceived enjoyment and attitude towards using. Therefore, system should be user-friendly to ensure the interface profile, logical clarity and color appearance so that students using the system can get more enjoyment of perception.
Using the AR system, students can combine the virtual environment with the real environment. This system can also reduce the cost of experiment and improve the safety of the experiment. Besides, it can also reduce the experimental space and increase the freedom of experiment compared with the traditional study. In addition, AR technology enables students to better situated learning, which can deepen the contact between learning content and real world.

This study investigates students’ technical acceptance degree and the determinants that influence students' attitudes. Future research can investigate using AR technology whether students actually acquire knowledge or compare them to traditional learning styles researching learning outcomes, satisfaction and other indicators.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (grant no 61602043).

References


An Empirical Case on Integration of Immersive Virtual Environment into Primary School Science Class

Jian SUNa, Hao LIa, Zhanhao LIUa, Su CAlab* & Xiaowen Lic

aSchool of Educational Technology, Faculty of Education, Beijing Normal University, China
bBeijing Advanced Innovation Center for Future Education, Beijing Normal University, China
cChina Mobile Communications Corporation Government and Enterprise Service Company, China
caisu@bnu.edu.cn

Abstract: Due to the technological advancement, the contemporary education is influenced to a great extent by high quality small displays, accelerometers and mobile devices. In this paper, two immersive virtual reality applications were developed and being applied to science classes in a primary school. During the in-field science classes teaching, students were divided into trios and they observed the resulting immersive VR applications by using VR Cardboard. With guidance of teachers, students were able to observe the virtual eclipse phenomenon from both ground and cosmic perspective staying in the classroom and thinking the reasons of eclipse. This research focuses on the effects of VR immersive applications on students’ knowledge achievements, new technology acceptability, learning behavior and satisfaction. From the questionnaire, it is found that learners have a high degree of acceptability for VR, and they are satisfied about learning in this way except for suffering motion sickness. In addition, VR applications are able to effectively and efficiently help students to understand and memory the knowledge of scientific phenomenon. While in the same time, the effect is slightly worse in terms of understanding the scientific principle.

Keywords: Virtual reality; integrating; immersion; science class

1. Introduction

Virtual Reality (VR), a computer simulation system that is used to create and experience the virtual world (Keppell & Macpherson, 1998), originated in the 1960s. After Facebook spent two billion dollars acquiring Oculus in 2014, virtual reality boomed again. Augmented Reality (AR) is the extension of Virtual Reality, which enables the combination of physical and virtual objects in a physical environment (Azuma et al., 2001). Year 2016 is called the first year of VR in China, when ‘VR heat’ took place in many areas including education. Chen (2006) asserts that “although VR is recognized as an impressive learning tool, there are still many issues needed further investigation, including: identifying the appropriate theories and/or models to guide its design and development”. Therefore, only rationally dealing with the relationship between technology and education can make educators better apply the technology to education.

More and more resources including time and capital have been being devoted to the designing and developing of desktop-based virtual reality instruction for teaching K-12 and higher education curriculum (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). Besides, some positive research outcomes about VR-based learning have been reported as enhanced anatomy learning (Petersson, Sinkvist, Wang, & Smedby, 2009); better performance in business knowledge application (Cheng & Wang, 2011); higher efficiency in matching diagrams and models (Stull, Barrett, & Hegarty, 2013) and so on. Hsien-Sheng (2016) with his partners has developed an augmented reality system called Weather Observers, which can help students learn about the atmospheric system and other geographical knowledge. For example, during the teaching session in the museum, students can scan the identification cards to get the model of the corresponding weather elements. The research shows that the augmented reality system has great advantages in motivating students' interest in geography and improving their learning effects. Cai, Chiang, Sun, Lin and Lee (2017) combined the AR with the Kinect somatosensory device to visualize the magnetic field. When learning about the knowledge of the magnetic field, students can interact with the device by hand gesture in order to understand the
distribution and change of it. Another research conducted by Cai’s team is about microscopic particle interaction experiment (Cai, Wang, & Chiang, 2014). They developed a chemistry learning tool based on augmented reality, which is intended to be introduced to middle schools. The research has shown that compared to traditional classes, teaching with AR based software can mobilize the enthusiasm of students, and make them more concentrated, so that students can perform better in remembering the structure of the atom. At the same time, the students are more impressed with what they have learned after they intuitively see the simulation models and interact with them. In addition, AR tools can improve students’ ability in experimental inquiry comparing to keyboard, mouse and computer operation. However, there are still some drawbacks of this tool. For example, the model is not exquisitely-designed, as well as the scene is not attractive enough.

On looking through the database, there are a few examples of integrating fully immersive virtual reality with science classes. This paper aims to design an immersive application of virtual reality based on the primary science curriculum standards. Then this application is used in classes and integration effect is verified. Some recommendations for the future work are also provided based on the experimental results.

2. Related Work

2.1 How to Integrate VR with Education

Virvou and Katsionis (2008) points out that “if games are to be introduced in classrooms, they have to be usable and likeable by the majority of students”. Similarly, virtual reality immersive application should also be ‘useable and likeable’ when introduced to classes. There is no “best way” to integrate technology into curriculum. Instead of that, integration efforts should be creatively designed and structured for particular subject matter ideas in specific classroom contexts (Koehler & Mishra, 2009). Therefore, when virtual reality is integrated with the class, educators should also pay attention to adjust the teaching structure in classroom flexibly. He (2014) believes that the “deep integration of technology and education” should achieve a structural transformation of the teaching system, which is, changing the role of teachers and students, so that students can become the leading part of classroom; change the teaching content and increase the computerized teaching resources; change the role of teaching media and make it an auxiliary learning tools. The Zone of Proximal Development (ZPD) theory proposed by Vygotsky (1964) requires that a learner should work in collaboration with a more capable partner, who is willing to provide appropriately challenging activities and a proper quantity and quality of assistance. And when integrating technology in classrooms, the role of teacher is supposed to be this more capable partner.

2.2 Three Categories of Virtual Reality

There are basically three different kinds of VR, categorized by the quality of the immersion provided by Cronin (1997), including “Non Immersive Systems”, “Semi- Immersive Immersive Systems” and “Fully Immersive Systems” (Kalawsky, 1996). Fully immersive VR is generally considering the best option for several reasons, including the ability to almost completely filter out interference from the outer reality world and thus allowing oneself to focus entirely on the virtual environment (Fällman, Backman, & Holmlund, 1999) And the study (Cronin, 1997) shows that immersive VR, where the learner is in a CAVE or wearing a head-mounted display, can be more efficient than monitor based desktop VR. Therefore, when doing experiments in classes, students are supposed to wear head-mounted display experiencing the full immersive VR.

2.3 Head-Mounted Displays used in Education

As VR technology is becoming more and more cost-effective, there are currently many Head-Mounted Displays (HMDs) in the market. Most of them have stereoscopic displays and tracking systems, enabling the user to see 3D images through a big field of vision and have the virtual camera moving according to the position of user’s head (Boas, 2013).

When creating an immersive environment in education, developers can use HTC Vive, Oculus and other professional heavyweight equipment as well as the lightweight equipment like Google
Cardboard. Take the Oculus as an example, it has been shown to present a significantly more realistic and compelling virtual world experience in comparison to traditional computer monitors (Reiners, Wood, & Gregory, 2014). There is a famous platform called “Unimersiv” (https://unimersiv.com/), which is one of the first VR-dedicated learning platforms available on Samsung’s Gear VR, the Oculus Rift, Day Dream and Google Cardboard. The developers publish new educational applications every month on the platform and the content concerns about history, space, human anatomy and etc. Some of the applications like “A Journey into the Human Brain” and “Explore the International Space Station” are graphically well designed and friendly to users.

Considering there is no enough space and high configuration computers in the classroom, it is decided to use the lightweight Head-Mounted Displays. Between two common lightweight displays, Google Cardboard is much easier to use than Gear and able to be matched with larger range of mobile phones, therefore, it is chosen to use displays like Google Cardboard in this study. However, people in PRC have no access to Google, so VR Cardboard which is nearly the same as Google Cardboard is used. In addition, although there are many wonderful VR applications on Google Play, they cannot be downloaded due to the network limitation. Also, the applications do not really match the teaching content of science education in primary school. Therefore, new immersive virtual reality applications were designed and developed to present an appropriate virtual learning environment.

3. Method

3.1 Development of Immersive Virtual Reality Application

Pantelidis (2010) proposes the 10-step model to determine when to use virtual reality in his study. According to this model and the benefits of virtual reality discussed by Mantovani (2001), “eclipse” was chosen as the course objective and two applications were designed. The development of immersive virtual reality application are consisted of five steps: selecting content, designing scene, designing user interface, piloting and modifying, and then being used in class (Figure 1). In order to help students observe a relatively lifelike phenomenon, the level of realism should be high. After the application was designed and built, it was tested and evaluated by a group of students from Beijing Normal University. Based on the evaluation results, the virtual environment was modified and the user-interface was improved.

![Figure 1. The development of VR immersive application](image-url)

3.1.1 Content Selection

Mastery of abstract scientific concepts require students to build flexible and runnable mental models (Redish, 1994). Frequently, these scientific models describe phenomena for which students have no real-life referents (Halloun & Hestenes, 1985). Students’ lack of real-life referents for intangible phenomena, coupled with an inability to reify abstract models, may result in their struggling with abstractions in science (Dede, Salzman, Loftin, & Sprague, 1999). Therefore, technology is supposed to aid students in experiencing the intangible phenomena in the classroom. Before the content selection, the sixth-year students were set as the experimental object. Then we looked through the content of the "Primary School Science Curriculum Standard", looking for content which is abstract or unobservable, and finally "eclipse" was selected to design the application.
3.1.2 The Basis of Designing

According to Piaget's division of children's cognitive development (Piaget, 2000) in four stages, sixth-grade students are in the Concrete Operations Stage where the feature of students’ thinking is that logical reasoning needs to rely on concrete image support. As a result, for the eclipse, we need to let students firstly have a specific perception in the virtual scene, and then let the students think about the causes of eclipse. In addition, according to Dede (2009), the more a virtual immersive experience is based on design strategies that combine dynamic, symbolic, and sensory factors, the greater the participant’s suspension of disbelief that she or he is “inside” a digitally enhanced setting. Therefore, when designing the virtual eclipse scene, all those domains should be taken into consideration.

3.1.3 The Scenes and Function of the Applications

The virtual reality immersive application designed for this paper is divided into two parts. The first part is about the walls of ancient Chinese city. Where the students can observe the eclipse phenomenon in virtual environment with the help of user interface. Students can also choose “playback” to observe again. The second part is observing the eclipse in a cosmic perspective. It can help students understand reasons of the eclipse by observing the trajectory of the sun, the moon and the earth in the universe. Specific content is shown in Figure 2.

![Figure 2](image)

Figure 2. Scenes - the eclipse occurs and comsim perspective

3.2 Design of the Four Elements in Class Structure

3.2.1 The Role of Teachers and Students

According to the Meaningful Learning Theory of Ausubel (1968), there are two ways to realizing meaningful learning. The first one is reception learning, which is a teacher–oriented teaching model, and the other discovery learning, is a student–oriented teaching model. Though the two models cast significant importance on the meaningful learning of students, both of them are not perfect with some minor defects. Based on these two teaching models, He (2007) puts forward Leader-Subject Instructional Structure, which underscores not only the leading role of teachers, but also the role of students as the subject of cognitions. Leader-Subject Instructional Structure was implemented in the process of immersive virtual reality-based instruction in this study, in which the teacher played the role as a director, and students were the subject of cognitions and learning activities.

3.2.2 Designing the Teaching Content

The content of eclipse, which is compulsory according to the curriculum standard, was chosen as the content of this experimental class. In the purpose of helping students get a better understanding of the eclipse, the class was consisted of several integrated parts from students’ perspective: watching a video of the eclipse playing on an electronic white-board, answering questions posted by the teacher, exploring the immersive virtual reality scene and taking part in group discussions.

3.2.3 Usage of Media

Teaching media was used for two reasons. First, when the teacher guided the students, an electronic
white-board was used to assist the teacher. Second, when students started to observe the scene and discuss their findings, the usage of VR Cardboard and the application purported to assist the students. With the assistance of media, the goal of the class was easier to be achieved.

3.3 Experiment

3.3.1 Participants and Procedure

Twenty-seven students from a primary school in Beijing District participated in this study. They were all in sixth year and had not taken any science class discussing the eclipse. None of the participants had prior exposure to VR Cardboard or immersive virtual reality applications. They were randomly divided into nine groups and were organized to take a two-period science class. Each group was given two sets of VR Cardboard and two mobile phones with the applications pre-installed.

During the first period, two videos introducing virtual reality technology were presented to students, and then students were asked to have an attempt on the VR Cardboards given to them. The purpose of the first period is to let students learn about and get familiar with virtual reality technology, prevent them from getting overexcited when using the immersive virtual reality application and thus get rid of factors interfering with the learning process and outcomes. During the second period, students were asked to observe the scenes presented by the application, discuss with their partners and answer questions posted by teachers, including explaining the cause, the whole process and the features of the eclipse. At the end of the class, all students were asked to fill in a questionnaire, including several questions about the eclipse and a scale testing the effect of applying virtual reality technology into science classes. The whole class was videotaped for further coding and analysis. Figure 3 shows the students experimenting with the application.

![Figure 3](image-url) The experiment in class - students observe the immersive VR APP by using VR Cardboard

3.3.2 Coding and Scale Analysis

This study used a coding system combining ITLAS Coding System by Jin and Gu (2010), and Teaching Media Coding System by Zhang and Wang. The system and the result is as below in Table 3.

The record was kept in a table in the format as Table 1 presents. According to rules of using ITLAS, coding was done every three seconds.

In order to test the effect of VR and the integration of primary school science curriculum, this paper designed a classroom integration effect test scale. Zhiye Li (2015) pointed out that in the evaluation of information technology and curriculum integration, it is necessary to consider not only the traditional evaluation of content, such as obtainment of knowledge and skills, but also the improvement in learning performance of students, such as learning outcomes, learning efficiency, learning methods, innovation ability and emotional attitude values. In addition, with the objects of applying virtual reality technology, a recent rise of new technology, to future classrooms, the degree of acceptability of the technology should also been taken into consideration.

This study divided the structure of VR and scientific classroom integration test into three parts: scientific knowledge, new technology acceptability, classroom performance and satisfaction. Science knowledge refers to the mastery of the knowledge about the eclipse. The new technology acceptability is based on the "Unified Theory of Acceptability and Use of Technology (UTAUT)" (Venkatesh & Davis, 2000) model, which includes four decisive factors – performance expectations, hard work expectations, social impact, and promotion conditions.

A total of twenty questions were set in three dimensions, as in Table 2. The first two questions are short answer questions. The rest are scale questions. The four choices of those questions are: “disagree” (1 point), “partly disagree” (2 points), “partly agree” (3 points) and “agree” (4 points).
The teacher of the science class and other teachers attending this class were given a short interview when the class was finished. They were all asked about opinions and suggestions on the integration of virtual reality technology into classroom.

Table 1: Sample of the coding record

<table>
<thead>
<tr>
<th>Starting time</th>
<th>Behavior code</th>
<th>Instructional Media code</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td>6</td>
<td>verbal</td>
<td>Introduction of the class</td>
</tr>
<tr>
<td>0:03</td>
<td>6</td>
<td>IL</td>
<td>Video playing</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 2: Structure and questions in the questionnaire

<table>
<thead>
<tr>
<th>Goal</th>
<th>Category</th>
<th>Questions</th>
</tr>
</thead>
</table>
|      | Mastery of knowledge | 1. The shadow first appears on which side of the moon observed from Earth?  
|      | Efforts expectation | 2. Please briefly describe the cause of the eclipse.  
|      | Performance expectation | 3. When using the VR application, I could easily control movements and observations in the scenes.  
|      | Classroom Performance and Satisfaction | 4. It didn’t take me a lot of effort to learn to use VR Cardboard and the VR application.  
|      |                      | 5. As for me, the content of the activities in the scenes are clear and easy to understand.  
|      |                      | 6. The content in the space scene helped me clearly understand the cause of the eclipse.  
|      |                      | 7. This application can be used everywhere  
|      |                      | 8. I easily managed to move in the scene and observe the eclipse  
|      |                      | 9. I felt dizziness or discomfort in my eyes.  
|      |                      | 10. I think using applications like this makes learning activities more plentiful.  
|      |                      | 11. Using applications like this is helpful when I am learning new knowledge.  
|      |                      | 12. The guidance provided by the VR application make my learning smoother.  
|      |                      | 13. Applications of this kind is helpful in improving my learning interest.  
|      |                      | 14. Applications like this makes learning easier.  
|      |                      | 15. Using VR applications made this class more interesting than the classes before.  
|      |                      | 16. I can discover new problems by learning with VR applications.  
|      |                      | 17. I think VR applications helped me to be more willing to cooperate with my classmates.  
|      |                      | 18. I like learning with VR.  
|      |                      | 19. I hope other subjects can also use VR applications.  
|      |                      | 20. I will recommend this way of learning to other fellow students.  

4. Results

4.1 Video Coding Analysis Results

The result of coding analysis of the video is presented as follows. In instructional media dimension, some types of media were not used in the class, thus they were not presented in Table 3. Based on the results, several indicators evaluating integration of technology in class were calculated. The results are as follows in Table 4.
Table 3: Coding record results

<table>
<thead>
<tr>
<th>Category</th>
<th>Content</th>
<th>Code</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher’s Language</td>
<td>Teacher’s acceptability of feeling</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Encouragement or praise from teacher</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Adoption of suggestions</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Posting open-ended question(s)</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Posting closed-ended question(s)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Direct Influence</td>
<td>Instruction</td>
<td>6</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>Direction</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Criticizing</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Students’ language</td>
<td>Answering passively</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Answering actively</td>
<td>10</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Asking questions</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Discussion with partners</td>
<td>12</td>
<td>108</td>
</tr>
<tr>
<td>Silence</td>
<td>Chaos in teaching</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Thinking of students</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Practice of students</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Instructional Media</td>
<td>traditional declarative media (blackboard)</td>
<td>TL</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>traditional interactive media</td>
<td>TI</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>technology assisted declarative media</td>
<td>IL</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>technology assisted interactive media</td>
<td>II</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>No media, only language is used</td>
<td>verbal</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 4: Coding record analysis results

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominance of class</td>
<td>Ratio of teacher’s speaking 29.87%</td>
</tr>
<tr>
<td></td>
<td>Ratio of students’ speaking 17.58%</td>
</tr>
<tr>
<td>Interaction behaviors between students and teachers</td>
<td>Ratio of indirect and direct influence 21.48%</td>
</tr>
<tr>
<td></td>
<td>Ratio of open-ended questions and closed-ended questions 660%</td>
</tr>
<tr>
<td></td>
<td>Ratio of student activeness 33.33%</td>
</tr>
<tr>
<td></td>
<td>Ratio of students’ discussion 59.02%</td>
</tr>
<tr>
<td>Appliance of technology</td>
<td>Ratio of technology usage 50.53%</td>
</tr>
<tr>
<td></td>
<td>Ratio of technology usage by teacher 14.63%</td>
</tr>
<tr>
<td></td>
<td>Ratio of technology by students 85.36%</td>
</tr>
</tbody>
</table>

Ratio of teacher’s speaking over students’ speaking is approximately 1.7:1, which means over a third of total speaking time is used by students. In addition, when students were using interactive media, the learning process was dominant by students. Teachers only walked around the classroom, giving instructions on operating of virtual reality devices occasionally. Thus, the instructional structure could be regarded as Leader-Subject Instructional Structure.

Ratio of indirect and direct influence was the result of dividing counts of indirect influence by direct influence. The ratio is 21.48%, which is below expectation, indicating that teacher mainly used instructional methods. However, the instructional methods were mainly used to give necessary instructions on operating mobile devices. Moreover, the teacher designed and asked much more open questions than closed questions (at a ratio of 33:5), indicating that there was abundant interaction among students and teachers.

The total time of technology using took up half of the class time. Students learnt knowledge of the eclipse mainly by interacting with instructional media.

### 4.2 Analysis of Scale

Each student was asked to fill in a questionnaire at the end of the class and all of the 27 questionnaires were collected and validity checked.
In the mastery of knowledge analysis, the correction rate of the first question was 100%. The correction rate of the second question was 88.8%. Among the answers, 11.2% didn’t answer, indicating they did not understand the cause of the eclipse. And 14.8% used graphs or language description to illustrate their ideas. From the result, conclusion can be concluded that the immersive virtual reality scenes were able to effectively help students understand and memory scientific phenomenon. The reason of this phenomenon is that further directions from the teacher need to be fully understood by all students.

Table 5: Scale results part I – Mastery of new Technology

<table>
<thead>
<tr>
<th>Question</th>
<th>Efforts Expectation</th>
<th>Performance Expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>3 4 5 6 7 8 9 10 11 12 13 14</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>3.67 3.81 3.67 3.37 3.52 2.59 3.89 3.81 3.74 3.81 3.67</td>
<td></td>
</tr>
<tr>
<td>St. d</td>
<td>0.54 0.47 0.47 0.54 0.67 0.63 1.04 0.31 0.39 0.44 0.47 0.47</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Scale results part II - Classroom Performance and Satisfaction

<table>
<thead>
<tr>
<th>Question</th>
<th>15 16 17 18 19 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>3.93 3.70 3.59 3.89 3.81 3.70</td>
</tr>
<tr>
<td>St. d</td>
<td>0.26 0.53 0.68 0.42 0.55 0.53</td>
</tr>
</tbody>
</table>

According to Table 5, five questions from the efforts expectation dimension scored above 3.5, indicating students tended to accept virtual reality technology, and were able to learn how to interact with the technology, including understanding the content of the scenes and controlling camera movements. In the performance expectation dimension, all questions scored above 3.5, indicating that immersive virtual reality technology were able to enrich students’ learning experience, increase their interest and help them get a better understanding of scientific knowledge.

Based on the two dimensions, students had high acceptability of immersive virtual reality applications and possessed strong willingness of using this technology.

According to Table 6, all questions in classroom performance and satisfaction dimension scored above 3.5, and the question “I found learning more fun when using virtual reality technology” had the highest score indicated that the most advantage of integrating virtual reality applications into science class was that it could effectively improve students’ interest on learning. Also, by integrating virtual reality technology into science classes, students’ creativity and cooperative ability can also be enhanced. In addition, students were eager to combine the using of virtual reality technology with learning, not only in science class, but also other subjects.

5. Conclusion & Discussion

From the instruction perspective, Leader-Subject Instructional Structure was successfully presented with the integration of virtual reality. From students’ perspective, they tended to accept the usage of virtual technology in their classes. Immersive virtual reality technology enriched students’ learning experience, increased their interest and helped them get a better understanding of scientific knowledge. In addition, students were eager to combine the using of virtual reality technology in learning, not only in science classes, but also in other subjects.

From the perspective of this experiment, some parts of the video were hard to code into one category, as several activities happened at the same time period. For example, when students were using the application, they might also be discussing with partners. This could result in an inaccuracy of the coding analysis. The coding system needs further development to be matched with different activities.

A criterion of the assessment of the scale needs to be further determined. The four choices of the question are “disagree” (1 point), “partly disagree” (2 points), “partly agree” (3 points) and “agree” (4 points). It can be seen that the value of the average scores indicates the opinions of most students in the class. Thus, a score above 3.5 means that most of the students agree with the statement. However, further research is needed to evaluate its effectiveness.
5.1 Advantages and Disadvantages of Integration of VR in Science Class

Virtual reality technology possesses strong expressive ability and sustainability. Compared with pictures and video instructional resources, it can provide more stimulation during study. Results of this study shows that the leading advantage of the integration of virtual reality into science classes is that it can inspire students’ interest. Virtual reality applications can make students easily immersed into scenes and exploring scientific knowledge. It also enriches the teaching methods in science classes. Students are willing to accept this learning method in terms of other subjects. Virtual reality applications also have advantages in improving learning efficiency and better understanding of knowledge.

Despite the advantages that VR possesses, there are still considerable obstacles towards widely integration of virtual reality into science classes. Unless a virtual reality application containing a systematic set of knowledge of science classes is developed, teachers with inadequate technology background would not be able to apply this technology to their classes.

Also, most of the students felt dizziness after using the application with VR Cardboard. The dizziness might be caused by several technical reasons. Fernandes and Feiner (2016) point out in their study that for moving users, high-quality tracking systems can minimize the mismatch between their visual perception of the virtual environment (VE) and the response of their vestibular system, diminishing VR sickness. For users who don’t move, by strategically and automatically manipulating field of view (FOV) during a VR session, the degree of participants’ VR sickness can be reduced, without decreasing their subjective level of presence.

Moreover, due to the limitations of development platforms, scientific details of the real world are difficult to be presented in the scenes. Some unreal details would make learners confused, which would not fully meet the requirement of education.

5.2 Suggestions

To solve the problems listed above, some suggestions are provided. First, an integrated, easy-to-use virtual reality instructional material developing platform need to be developed. For this platform, teachers should be able to develop and modify their own scenes based on given models and materials. These models and materials needed to be scientific and close to real world. Therefore, to take advantage of applying virtual reality to classes, a simple, easy-to-use and scientific platform is required. Second, a new instruction environment and new learning activities need to be designed for the integration of immersive virtual reality into class, to avoid motion sickness. Students’ experience using the technology would be continually improved. When using VR application in classes, educators can design some activities to ensure students’ actual movements matching their visual perception.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No 61602043).

References


Online Synchronous Discussion in Face-to-Face Classroom Based on WeChat

Manli LUa, Guang CHEN ab & Shuxian OUYANGa

Abstract: In recent years, the online synchronous classroom has developed rapidly in higher education, which has attracted the attention of many educational scholars. This study collected online synchronous discussion records during two months based on WeChat of an undergraduate students group and a graduate students group. The undergraduate group contains 355 messages and the graduate group contains 521 messages. All messages were coded by two trained coders. Content analysis and lag sequential analysis were conducted to explore the behavioral patterns. The analyses also revealed that the characteristics and differences of the interaction behavior discussed by different learners. Undergraduate students and graduate students are good at proving the rationality of the view. They have a lot of the behavior of emotional communication in the classroom. Graduate students are more adept at asking questions and thinking deeply than undergraduate students. And graduate students are better at crossing three dimensions of behavior than undergraduate students. Instructional suggestions were proposed to facilitate further online synchronous classroom interaction.

Keywords: WeChat, blended learning, learning behavioral pattern, online discussion activity, instant messaging tools

1. Introduction

With the rapid development of Internet, the online education has become another development field of helping education (Amador & Amador, 2014). The popularity of instant messenger leads to the emergence of a new instruction model. Online synchronous learning has become increasingly popular because of its conveniently and many studies have proven the benefits of online synchronous teaching and learning (Chen, Ko, Kinshuk, & Lin, 2005). The emergence of MOOC (Massive Open Online Course) makes it possible to learn anytime and anywhere. The popularity and application of smart phones affect learners' learning habits. The emergence of mobile learning allows learners to acquire educational information, educational resources, and educational services by using wireless mobile technology (Briz-Ponce, Pereira, Carvalho, Juanes-Méndez, & García-Peñalvo, 2017). This is a profound impact on traditional classroom teaching and it also promotes the reform of classroom teaching. Blended learning combines the advantages of traditional methods of teaching and learning and online learning (W. Chen & Looi, 2007a). In this process, the learner's classroom behavior pattern will also change, which from passive acceptance the knowledge taught by teachers into a more active and initiative to think and create.

2. Literature Review

Discussion has played a vital role in teaching and learning. Studies have found out that classroom discussion was significantly related to learning performance (Apple-bee, Langer, Nystrand, & Gamoran, 2003). A number of studies have shown that students behave differently between face-to-face and online discussion, for example, students who are less active in face-to-face discussion were more active online and some students argued that they were less stressful in an online learning environment than the traditional classroom (Kelm, 1992; Kern, 1995). With the increasing popularity of online
learning, a growing body of literature has explored the characteristics of online discussion. Online
discussion can be categorized to online asynchronous discussion and online synchronous discussion.
Literatures about online asynchronous discussion. The self-regulated, self-paced quality was one of the
most attractive features of online discussion (Tiene, 2000). And the written record was good for review
and reflection.

Hsieh & Tsai (2012) explored the role of moderators in online synchronous discussion. The
results showed that the moderator helped the students to enhance their collaboration pattern and to
increase the online participation rate. The strategies of helping students focus on the main topic and
giving students feedback were also crucial for online synchronous discussion.

Previous study compared the in-class and off-class online discussion and found out that in-class
postings were less interactive than off-class ones. The main reason of this finding was the lack of
interaction perceived by the learners (Chen & Looi, 2007b). Some researchers argued that online
discussion has disadvantages like technical barriers and a lack of visual cues. Educators use lots of tools
to facilitate face-to-face and online discussion. Some of them are specific tools, such as Instant
Response Systems. Some of them are not. Social medias, such as Facebook, twitter, etc., are used as
efficient tools. However, current research primarily focuses on traditional classroom and online
asynchronous discussion. Previous research has seldom discussed online synchronous discussion in
traditional face-to-face classroom settings. WeChat as a kind of instant messaging tool, the study online
synchronization discussion mode of undergraduate students and graduate students in the WeChat is
very meaningful and needed.

Accordingly, the present study conducts an online synchronous discussion in face-to-face
classroom. The research questions of this study are as follows.

Q1: what’s the difference of typical online synchronous discussion activities between
undergraduate students and graduate students?
Q2: what’s the difference of online synchronous discussion patterns between undergraduate
students and graduate students?

3. Research Methods

3.1 Participants

Two blended learning courses were selected in the same school year. Students took part in these two
courses were asked to carry out a series of learning tasks including online synchronous discussion on an
instant messenger app called WeChat. WeChat is the most popular instant messenger app in China. The
two courses were fundamental psychology courses and they were taught by a same instructor. The
instructor carried out a similar instructional approach combining traditional lecturing and online
discussion in both courses. All participants were learning these courses for the first time. The
undergraduate group included 48 freshmen from Education major and the course is "Introduction to
Psychology". The graduate group included 67 first year students from Education major and the course is
"Developmental Psychology". Participants were asked to join the WeChat group at the beginning of
each course. They can communicate with instructor and other students via text, voice, and pictures
messages during the face-to-face lecturing sessions. All participants are familiar with the function of
WeChat.

3.2 Procedure

This study lasted for two months. WeChat was used as an assist teaching tool. Students were allowed to
post anything in the discussion group. The instructor would not force the students to participate in the
discussion and the discussion would not affect the final grade. At the end of the research session, the
discussion log was exported for subsequent analysis.
3.3 Coding Scheme

Researchers found that there were three different behaviors in online course, including cognitive presence, social presence, and teaching presence (Anderson, Rourke, Garrison, & Archer, 2001). Cognitive presence can be divided into five categories: sharing, demonstration, negotiation, produce, and reflection (Liu, Zhu, Chen, & Huang, 2005). Some researchers believed that social presence includes positive emotions, negative emotions, seeking help, asking questions, and explanation or providing help (Zhu, Liu, & Huang, 2007). Anderson et al (2001) also claimed that teaching presence can be divided into design and organization, facilitative discussion, and direct instruction. Therefore, the coding scheme used in the present study was adopted from the above-mentioned studies which included cognitive, social, and teaching dimensions (as shown in Table 1).

Table 1: Coding scheme for analyzing undergraduate and graduate behavior.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Category</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Presence</td>
<td>Share (S)</td>
<td>Supply or introduce relative fact, resource, information or knowledge.</td>
</tr>
<tr>
<td></td>
<td>Demonstrate (D)</td>
<td>Judge, comment, demonstrate, explain and summarize the conclusion, viewpoint, fact information.</td>
</tr>
<tr>
<td></td>
<td>Negotiate (N)</td>
<td>Check, affirm, doubt the viewpoint; agree or disagree and modify the viewpoint.</td>
</tr>
<tr>
<td></td>
<td>Produce (P)</td>
<td>Synthesize the views or statements, make refinement and summary, and thus sum up the conclusions or products.</td>
</tr>
<tr>
<td></td>
<td>Reflect (R)</td>
<td>Reflect the learning process, methods and achievement.</td>
</tr>
<tr>
<td>Social Presence</td>
<td>Positive emotion (PE)</td>
<td>Express or describe personal positive emotion.</td>
</tr>
<tr>
<td></td>
<td>Negative emotion (NE)</td>
<td>Express or describe personal negative emotion.</td>
</tr>
<tr>
<td></td>
<td>Seek help or ask a question (HE)</td>
<td>Ask others for help or help others, state something irrelevant with academic tasks.</td>
</tr>
<tr>
<td></td>
<td>Explain or provide the help (EE)</td>
<td>Answer other people's help, state their own status or explain things related to the task submission.</td>
</tr>
<tr>
<td>Teaching Presence</td>
<td>Organizing teaching (O)</td>
<td>Design activities, determine the time to complete the task and the media, establishing ritual constraints.</td>
</tr>
<tr>
<td></td>
<td>Facilitate discussion (F)</td>
<td>Inspire and guide a talk by asking questions, act as organizer and guide the participation.</td>
</tr>
<tr>
<td></td>
<td>Lead thinking (L)</td>
<td>Teachers provide intellectual and scholarly leadership and share their subject matter knowledge with students.</td>
</tr>
<tr>
<td>Others</td>
<td>Others (OS)</td>
<td>Any behaviors not mentioned above.</td>
</tr>
</tbody>
</table>

Two trained coders coded the discussion log. If one message was divided into two or more messages, they would be coded as one single message. In addition, the discussion log contained different types of messages, including text, voice, pictures, web links, and videos. The coder took them into account and coded them too. The inter-coder Kappa coefficient of undergraduate group was 0.872 and the inter-coder Kappa coefficient of graduate group was 0.907. Finally, total number of 355 messages from undergraduate group and 521 messages from graduate group were coded.

4. Results

4.1 Content Analysis

Content analysis has been used by many researchers to study learners’ behavior patterns (Hou, Sung, & Chang, 2009; Cheng & Hou, 2015). The content analysis includes both qualitative content analysis and quantitative content analysis. Quantitative content analysis has been used in several studies to explore the online discussion process (Jeong, 2003). Researchers believed that coding scheme was the most important part of content analysis (Garrison, Anderson, & Archer, 1999).
The frequencies of messages from two groups (as shown in Table 2 and Figure 1) were coded according to coding scheme (see Table 1).

### Table 2: Frequencies of undergraduate and graduate behavior.

<table>
<thead>
<tr>
<th>Category</th>
<th>Undergraduate Group</th>
<th>Graduate Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>Cognitive presence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>55</td>
<td>15.5%</td>
</tr>
<tr>
<td>D</td>
<td>36</td>
<td>10.1%</td>
</tr>
<tr>
<td>N</td>
<td>50</td>
<td>14.1%</td>
</tr>
<tr>
<td>P</td>
<td>15</td>
<td>4.2%</td>
</tr>
<tr>
<td>R</td>
<td>7</td>
<td>2.0%</td>
</tr>
<tr>
<td>total</td>
<td>163</td>
<td>45.9%</td>
</tr>
<tr>
<td>Social presence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>72</td>
<td>20.3%</td>
</tr>
<tr>
<td>NE</td>
<td>30</td>
<td>8.5%</td>
</tr>
<tr>
<td>HE</td>
<td>25</td>
<td>7.0%</td>
</tr>
<tr>
<td>EE</td>
<td>35</td>
<td>9.9%</td>
</tr>
<tr>
<td>total</td>
<td>162</td>
<td>45.6%</td>
</tr>
<tr>
<td>Teaching presence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>3</td>
<td>0.8%</td>
</tr>
<tr>
<td>F</td>
<td>4</td>
<td>1.1%</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>total</td>
<td>8</td>
<td>2.3%</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>22</td>
<td>6.2%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>355</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

**Figure 1.** Distribution of content analysis of undergraduate and graduate behavior.

As shown in Table 1, PE (positive emotions) accounted for the largest proportion in undergraduate group and graduate group. For undergraduate students, cognitive presence and social presence accounted for almost equal proportion. The percentage were 45.9% and 45.6% respectively. The proportion of S (share) in cognitive presence was 15.5%, followed by N (negotiation) accounted for 14.1%. PE (positive emotions) in the social presence accounted for a large proportion of 20.3%. The teaching presence was less frequent and 2.3% less. Other discussions also accounted for a certain proportion of 6.2%.

For graduate students, social share accounted for 74.3%, cognitive presence accounted for only 16.3%. N (negotiation) in cognitive presence accounted for a higher proportion of 5.2%, followed by S (shared) accounted for 4.8%. The proportion of PE (positive emotions) in the Social presence accounted for a larger 29.4%, HE (Help or Questions) and EE (Explain or Offer the Help) appeared a higher frequency, the proportion accounted for 18.6% and 17.9%. Teaching presence and the other accounted for less, 4.4% and 5.0% respectively.

The distribution showed that interactions occurred in online synchronous discussion in face-to-face classroom were mostly composed of knowledge sharing, negotiation, expressing positive...
feelings, asking questions and asking for help, and explaining or providing help. Most of the interaction between teachers and students occurred in both cognitive and social dimensions. It can be known that the participant's attention was high and the participant could effectively receive the teacher's information and give a timely response in online synchronous discussion in face-to-face classroom. In addition, the emotional communication between teachers and students was more frequent. There was emotional exchange among participants before entering formal teaching and after the teaching activities, which will help create a relaxed and friendly teaching and learning atmosphere. However, the frequency of P (produce) and R (reflection) was less in interactive behavior. This shows that the interaction among participants was more simple knowledge sharing and negotiation, less of the knowledge of the summary and reflection.

The proportion of teaching presence involved teachers and teaching assistants was less. O (organizing teaching), F (facilitate discussion), L (lead thinking) accounted for a small proportion. The exchange and feedback between teachers and students was important classroom activities in online synchronous discussion in face-to-face classroom.

The chi-square test was performed on the distribution of the interaction between undergraduate group and graduate group (P = 0.000 <0.05). It indicated that the interaction behavior of the two groups was significantly different from each other. More interactive behavior took place in positive emotions in the two groups. During the two months’ period, the graduate students sent more messages than the undergraduate students. From the social interaction point of view, however, it can be found that undergraduate students were more willing to share ideas and consult with others, graduate students were more willing to express emotions, questions and explanations. The questions raised by undergraduates are more likely to be answered by teachers, and the questions raised by the graduate students are answered by classmates, and the dependency of the graduate students on the teachers was relatively small. P (produce) and R (reflection) are rare for undergraduate students and graduate students. Most of the ideas are summarized by the teacher, the reflection was done by the students, which shows that the students in the discussion are more accustomed to express their views through the demonstration and negotiation, not good at summing up the views of the conclusions.

4.2 Behavioral Sequence Analysis (LSA)

The lag sequential analysis (Bakeman & Gottman, 1997) is a way to explain the sequential behavioral patterns based on discussion log. The main function of lag sequential analysis is to observe the sequence of behavior that continuously occurs at different stages of online discussion. That is, to observe the next situation and the total number of behavior after the end of a particular discussion stage. Lag sequential analysis (Bakeman & Gottman, 1997; Hou, 2010) enables us to conduct a more structured and visualized analysis of behavioral patterns. The data is analyzed by time sequence. The significant behavioral sequences compile an overall pattern of behavior in discussion (Hou, Sung, & Chang, 2009).

The study used the lag sequential analysis to analyze the behavior patterns of the participants. First, the coded data table was converted into the behavior sequence frequency table. Second, the behavior sequence frequency table was converted to the adjustment residual Z-score table. The column represents the previous behavior and the row represents the latter behavior. This Z-score is revised according to a certain formula, so it is called the adjustment residual Z-score. It is generally believed that the Z-score greater than 1.96 is to achieve statistical significance (p <0.05). Through the lag sequential analysis, it can be found there were 18 significant behavioral sequences in the undergraduate group and 23 significant behavioral sequences in the graduate group.

Table 3 presents Z-scores of online synchronous discussion behavior of the undergraduate group. The significant behavioral sequences were as follows: S→S (4.23), S→D (4.08), N→D (2.48), N→N (4.35), N→P (3.70), P→N (4.46), R→N (2.20), R→R (2.36), PE→PE (5.13), PE→O (2.00), NE→NE (9.91), HE→EE (5.23), HE→L (3.63), EE→EE (3.90), O→HE (6.31), F→F (4.54), L→EE (3.02), OS→OS (6.96).

Table 4 presents Z-scores of online synchronous discussion behavior of graduate group. The significant behavioral sequences were as follows: S→D (10.70), S→N (2.50), D→N (9.21), D→P (2.50), N→S (5.42), N→D (5.10), N→N (4.10), P→S (5.07), P→P (3.74), P→R (4.72), R→NE (2.11), PE→PE (5.77), PE→F (2.02), NE→NE (4.12), HE→EE (6.07), HE→O (1.97), EE→HE (2.54), O→O (2.22), O→L (3.04), L→S (2.38), OS→R (2.93), OS→L (2.26), OS→OS (6.96).
Table 3: Z scores of online synchronous discussion behavior of undergraduate group.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>D</th>
<th>N</th>
<th>P</th>
<th>R</th>
<th>PE</th>
<th>NE</th>
<th>HE</th>
<th>EE</th>
<th>O</th>
<th>F</th>
<th>L</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>4.23</td>
<td>4.08</td>
<td>0.10</td>
<td>0.49</td>
<td>0.96</td>
<td>-2.94</td>
<td>-1.40</td>
<td>-1.65</td>
<td>-1.69</td>
<td>-0.75</td>
<td>0.53</td>
<td>-0.43</td>
<td>-2.08</td>
</tr>
<tr>
<td>D</td>
<td>1.75</td>
<td>-0.92</td>
<td>0.03</td>
<td>1.34</td>
<td>0.39</td>
<td>0.88</td>
<td>-1.90</td>
<td>0.37</td>
<td>1.47</td>
<td>1.37</td>
<td>-0.67</td>
<td>-0.33</td>
<td>-0.87</td>
</tr>
<tr>
<td>N</td>
<td>0.10</td>
<td>2.48</td>
<td>4.35</td>
<td>3.70</td>
<td>0.01</td>
<td>-2.30</td>
<td>-2.32</td>
<td>-1.51</td>
<td>-2.02</td>
<td>-0.71</td>
<td>-0.82</td>
<td>-0.41</td>
<td>-1.33</td>
</tr>
<tr>
<td>P</td>
<td>-0.24</td>
<td>0.41</td>
<td>4.46</td>
<td>-0.83</td>
<td>1.33</td>
<td>-1.32</td>
<td>-1.20</td>
<td>-0.06</td>
<td>-1.31</td>
<td>-0.37</td>
<td>-0.42</td>
<td>-0.21</td>
<td>-1.02</td>
</tr>
<tr>
<td>R</td>
<td>-0.09</td>
<td>0.36</td>
<td>2.20</td>
<td>-0.56</td>
<td>2.36</td>
<td>-0.39</td>
<td>-0.81</td>
<td>-0.74</td>
<td>-0.89</td>
<td>-0.25</td>
<td>-0.29</td>
<td>-0.14</td>
<td>-0.69</td>
</tr>
<tr>
<td>PE</td>
<td>-1.16</td>
<td>-1.45</td>
<td>-2.34</td>
<td>-1.34</td>
<td>-1.35</td>
<td>5.13</td>
<td>-0.52</td>
<td>0.47</td>
<td>-0.94</td>
<td>2.00</td>
<td>0.23</td>
<td>-0.51</td>
<td>0.83</td>
</tr>
<tr>
<td>NE</td>
<td>-0.88</td>
<td>-0.66</td>
<td>-2.32</td>
<td>-1.20</td>
<td>-0.81</td>
<td>-0.48</td>
<td>9.91</td>
<td>-1.58</td>
<td>0.02</td>
<td>-0.53</td>
<td>-0.61</td>
<td>-0.30</td>
<td>-1.47</td>
</tr>
<tr>
<td>HE</td>
<td>-1.65</td>
<td>-1.06</td>
<td>-1.51</td>
<td>-1.09</td>
<td>-0.74</td>
<td>-1.04</td>
<td>-0.09</td>
<td>1.00</td>
<td>5.23</td>
<td>-0.48</td>
<td>-0.55</td>
<td>3.63</td>
<td>1.24</td>
</tr>
<tr>
<td>EE</td>
<td>-1.69</td>
<td>-2.10</td>
<td>-0.99</td>
<td>-0.43</td>
<td>0.39</td>
<td>1.33</td>
<td>0.02</td>
<td>-0.33</td>
<td>3.90</td>
<td>-0.58</td>
<td>1.02</td>
<td>-0.33</td>
<td>-0.13</td>
</tr>
<tr>
<td>O</td>
<td>-0.75</td>
<td>-0.59</td>
<td>-0.71</td>
<td>-0.37</td>
<td>-0.25</td>
<td>-0.87</td>
<td>-0.53</td>
<td>6.31</td>
<td>-0.58</td>
<td>-0.16</td>
<td>-0.19</td>
<td>-0.09</td>
<td>-0.45</td>
</tr>
<tr>
<td>F</td>
<td>0.53</td>
<td>-0.68</td>
<td>-0.82</td>
<td>-0.42</td>
<td>-0.29</td>
<td>0.25</td>
<td>-0.61</td>
<td>1.41</td>
<td>-0.67</td>
<td>-0.19</td>
<td>4.54</td>
<td>-0.11</td>
<td>-0.52</td>
</tr>
<tr>
<td>L</td>
<td>-0.43</td>
<td>-0.34</td>
<td>-0.41</td>
<td>-0.21</td>
<td>-0.14</td>
<td>-0.50</td>
<td>-0.30</td>
<td>-0.28</td>
<td>3.02</td>
<td>-0.09</td>
<td>-0.11</td>
<td>-0.05</td>
<td>-0.26</td>
</tr>
<tr>
<td>OS</td>
<td>-1.47</td>
<td>-1.63</td>
<td>-1.33</td>
<td>-1.02</td>
<td>-0.69</td>
<td>-0.23</td>
<td>-0.68</td>
<td>2.10</td>
<td>-0.13</td>
<td>-0.45</td>
<td>-0.52</td>
<td>-0.26</td>
<td>6.96</td>
</tr>
</tbody>
</table>

Table 4: Z scores of online synchronous discussion behavior of graduate group.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>D</th>
<th>N</th>
<th>P</th>
<th>R</th>
<th>PE</th>
<th>NE</th>
<th>HE</th>
<th>EE</th>
<th>O</th>
<th>F</th>
<th>L</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>1.80</td>
<td>10.70</td>
<td>2.50</td>
<td>0.67</td>
<td>-0.32</td>
<td>-1.06</td>
<td>-1.56</td>
<td>-2.45</td>
<td>-1.86</td>
<td>-0.91</td>
<td>-0.45</td>
<td>-0.39</td>
<td>-1.18</td>
</tr>
<tr>
<td>D</td>
<td>1.17</td>
<td>0.27</td>
<td>9.21</td>
<td>2.50</td>
<td>-0.28</td>
<td>-1.94</td>
<td>-1.39</td>
<td>-1.01</td>
<td>-1.53</td>
<td>-0.81</td>
<td>-0.40</td>
<td>-0.35</td>
<td>-1.05</td>
</tr>
<tr>
<td>N</td>
<td>5.42</td>
<td>5.10</td>
<td>4.10</td>
<td>1.96</td>
<td>-0.33</td>
<td>-3.01</td>
<td>-0.91</td>
<td>-2.05</td>
<td>-1.97</td>
<td>0.19</td>
<td>-0.47</td>
<td>-0.41</td>
<td>-0.32</td>
</tr>
<tr>
<td>P</td>
<td>5.07</td>
<td>-0.67</td>
<td>1.96</td>
<td>3.74</td>
<td>4.72</td>
<td>-1.50</td>
<td>-1.02</td>
<td>-1.61</td>
<td>-1.56</td>
<td>-0.60</td>
<td>-0.30</td>
<td>-0.26</td>
<td>0.63</td>
</tr>
<tr>
<td>R</td>
<td>-0.31</td>
<td>-0.28</td>
<td>-0.33</td>
<td>-0.21</td>
<td>-0.09</td>
<td>0.64</td>
<td>2.11</td>
<td>-0.68</td>
<td>-0.66</td>
<td>-0.25</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.33</td>
</tr>
<tr>
<td>PE</td>
<td>-1.85</td>
<td>-2.93</td>
<td>-2.56</td>
<td>-0.81</td>
<td>-0.91</td>
<td>5.77</td>
<td>-0.99</td>
<td>-0.34</td>
<td>-2.06</td>
<td>-0.94</td>
<td>2.02</td>
<td>0.16</td>
<td>1.06</td>
</tr>
<tr>
<td>NE</td>
<td>-0.79</td>
<td>-1.39</td>
<td>-1.62</td>
<td>-1.02</td>
<td>-0.43</td>
<td>-0.33</td>
<td>4.12</td>
<td>0.32</td>
<td>0.46</td>
<td>-1.24</td>
<td>1.19</td>
<td>-0.53</td>
<td>-0.87</td>
</tr>
<tr>
<td>HE</td>
<td>-2.34</td>
<td>-2.13</td>
<td>-2.49</td>
<td>0.03</td>
<td>-0.66</td>
<td>0.41</td>
<td>0.88</td>
<td>2.54</td>
<td>1.01</td>
<td>0.75</td>
<td>-0.94</td>
<td>-0.81</td>
<td>-1.39</td>
</tr>
<tr>
<td>EE</td>
<td>-0.89</td>
<td>0.51</td>
<td>-0.95</td>
<td>-0.60</td>
<td>-0.25</td>
<td>0.72</td>
<td>-0.32</td>
<td>0.66</td>
<td>-1.23</td>
<td>2.22</td>
<td>-0.36</td>
<td>3.04</td>
<td>-0.93</td>
</tr>
<tr>
<td>O</td>
<td>-0.44</td>
<td>-0.40</td>
<td>1.79</td>
<td>-0.30</td>
<td>-0.12</td>
<td>-1.30</td>
<td>-0.61</td>
<td>1.62</td>
<td>-0.94</td>
<td>-0.36</td>
<td>-0.18</td>
<td>-0.15</td>
<td>1.84</td>
</tr>
<tr>
<td>L</td>
<td>2.38</td>
<td>-0.35</td>
<td>-0.41</td>
<td>-0.26</td>
<td>-0.11</td>
<td>0.15</td>
<td>-0.53</td>
<td>0.65</td>
<td>-0.81</td>
<td>-0.31</td>
<td>-0.15</td>
<td>-0.13</td>
<td>-0.40</td>
</tr>
<tr>
<td>OS</td>
<td>-1.15</td>
<td>-1.05</td>
<td>-1.22</td>
<td>-0.77</td>
<td>2.93</td>
<td>0.15</td>
<td>-0.14</td>
<td>0.08</td>
<td>-0.87</td>
<td>-0.93</td>
<td>-0.46</td>
<td>2.26</td>
<td>4.34</td>
</tr>
</tbody>
</table>

According to the Z-scores, transition diagrams of online synchronous discussion of both groups were shown in Figure 2 (undergraduate group) and Figure 3 (graduate group).
The values in the figure represent the Z values of each sequence, the thickness of the arrow indicates a significant degree, and the behavior is shifted in the direction of the arrow. These sequences showed the overall interaction behavior patterns of the two groups in online synchronous discussion.

4.2.1 The Sequential Patterns of undergraduate group

By analyzing the overall interaction of undergraduate students, it can be observed that the behavioral sequences of the six sequences were significantly higher, including NE→NE (9.91), OS→OS (6.96), PE→PE (5.13), F→F (4.54), S→S (4.23), N→N (4.35). However, the significance of P (produce) was still small. This showed that six kinds of behavior patterns including share, negotiation, positive emotion, negative emotion, lead thinking and others were the ongoing classroom undergraduate’s activities in online synchronous discussion in face-to-face classroom.

The first was the pattern of behavior that occurs between reflection, negotiation, produce, demonstration of cognitive presence dimensions. It was R→N→P and R→N→D. N (negotiate) was usually followed after R (reflect). N (negotiate) would lead to different behavior patterns, including P (produce) and D (demonstrate). This suggested that once there is little frequency of reflection, it would cause the learner's attention. Then the learners to negotiate and discuss. At this time there may be two cases. One was the teacher or the learner to sum up the conclusion, the other was the learners continue to clarify their points. This behavior pattern promoted the behavior of cognitive dimensions.
The second was the pattern of behavior that occurs between teaching presence and social presence. It was \( \text{O} \rightarrow \text{HE} \rightarrow \text{EE} \) and \( \text{HE} \rightarrow \text{L} \rightarrow \text{EE} \). The occurrence of a teacher's teaching organization (O) would cause the learner to ask for help or ask questions (HE), then someone would explain or provide help (EE). When a learner raised a question, teachers or assistants usually lead the students to think (L), and then someone would propose a solution. This suggested that there was cross between teaching presence and social presence. After the teacher consciously organized the teaching, the students would put forward questions, and feedback on the questions was also very timely. Then after students ask questions, teachers would appear to lead thinking behavior to encourage students to think deeply and create a solution. The emergence of this positive behavior model facilitated the cross of teaching presence and social presence.

The third was the pattern of behavior that occurs between sharing and demonstration. It was \( \text{S} \rightarrow \text{D} \). Students liked to share their views in the classroom. And then they would prove the rationality of their own point of view or other students to further analyze this view.

4.2.2 The Sequential Patterns of Graduate Group

By analyzing the overall interaction of graduate students, it can be observed that the behavioral sequences of the six sequences were significantly higher, including \( \text{OS} \rightarrow \text{OS} (6.96) \), \( \text{PE} \rightarrow \text{PE} (5.77) \), \( \text{NE} \rightarrow \text{NE} (4.12) \), \( \text{N} \rightarrow \text{N} (4.10) \), \( \text{P} \rightarrow \text{P} (3.74) \), \( \text{O} \rightarrow \text{O} (2.22) \). This showed that six kinds of behavior patterns including negotiation, produce, positive emotion, negative emotion, organizing teaching and others are the ongoing classroom graduate’s activities in online synchronous discussion in face-to-face classroom. Participants have a high attention and stamina in classroom activities.

In this interactive behavior pattern, four independent and significant behavior pattern sequences were found.

The first was the behavior pattern of the cycle that occurs between sharing, demonstration, negotiation of cognitive presence dimensions. It was \( \text{S} \rightarrow \text{D} \rightarrow \text{N} \rightarrow \text{S} \). D (demonstration) is usually followed after S (share). Demonstration (D) would lead to the emergence of N (negotiation). After the discussion of the learners, the learners would usually put forward their new ideas (S). This suggested that learners will demonstrate after expressing their views. Then other learners would discuss, which led to the creation and discussion of new ideas. The behavioral pattern of this cycle would make the discussion behavior more sustainable and promote the discussion of cognitive dimensions.

The second was the behavior pattern that occurs between sharing, produce and reflection of cognitive presence dimensions. It was \( \text{P} \rightarrow \text{S}, \text{P} \rightarrow \text{R} \). After the learners synthesize the views and form the conclusions, there are usually other students share new insights or learners to reflect on interactions. The patterns of behavior that occur at the cognitive dimension and social dimension are \( \text{P} \rightarrow \text{R} \rightarrow \text{NE} \). This showed that learners may have negative emotions after concluding and reflecting. This is because after the learners reflect on the events, words and deeds, questions, experiences in the interaction, they will find their own problems and shortcomings, so they may show negative emotions such as anxiety, dissatisfaction and inferiority.

The third was the pattern of behavior that occurs between sharing, produce and reflection of social presence dimensions. This indicated that other students provide helping after one raise the question. After solving the problem, there would be more people to raise new questions. Learners do depth thinking about the problem.

The fourth was the pattern of behavior that occurs between teaching presence and cognition presence. It was \( \text{O} \rightarrow \text{L} \rightarrow \text{S} \). After the organization of teaching, teachers usually take action to continue to promote the discussion of learners, then learners will put forward their views. This showed that teachers would consciously organize teaching and encourage students to discuss to enable students to share their views.

4.2.3 The Difference of Sequential Patterns between Undergraduate and Graduate Group

The behavioral patterns in cognitive dimensions of undergraduates are relatively independent. When they were in occurrence of cognitive behavior, the undergraduate students won't appear the behavior of the other dimensions. However, when they were in occurrence of cognitive interactive behavior, graduates’ behavior was to facilitate discussion of teaching presence dimensions and negative emotion
of social presence dimensions. This showed that the graduate students are more adept at making the behavior of the three dimensions cross occurs than undergraduate students.

At the social presence dimension, both undergraduate students and graduate students have a pattern of behavior was asking questions and then solving problems. The difference was that undergraduates’ behavior patterns are usually teachers first organize teaching, and then students ask questions, and finally others to solve the problem. However, graduate students usually continue to ask new questions after solving the problem. This is because the graduate students are more adept at asking questions and thinking deeply.

At the teaching presence dimension, the same between undergraduate students and graduate students was the emergence of the behavior of teaching dimensions, which lead to the occurrence of social dimension. The difference was that teachers have more guide thinking behavior in the undergraduate group. Teachers have more organization teaching behaviors in the graduate group. This is because the graduates would spontaneously carry out depth thinking, undergraduates lack of consciousness, so the teacher's guidance is necessary.

Other discussions that are unrelated to the teaching content was common classroom activity that occur in undergraduate and graduate classes. The difference was that the other discussions in the undergraduate classroom are relatively independent and do not lead to other behaviors, but other discussions in the graduate's class will lead to the teacher's promotion of discussion and student reflection. This indicated that undergraduates’ unrelated discussions were less affected by other behaviors or affect other behaviors, but the teacher would guide the graduate students to return to the classroom or to self-reflection through the organization of teaching after the unrelated behavior of the graduate students.

5. Conclusion

This study was about online synchronous discussion in face-to-face classroom based on WeChat, and encoded the discussion messages of undergraduate students and graduate students. Then participants' interactive behavior patterns were found by content analysis and lag sequential analysis. And studied the characteristics and differences of the interactive behavior patterns between undergraduate students and graduate students. Positive emotional communication was the most important activities in effective discussing both undergraduate students and graduate students.

It can be found by quantitative content analysis and lag sequential analysis as follows. In the undergraduate classroom, the frequency of positive emotional communication, knowledge sharing, negotiation was higher and was significantly higher classroom behaviors. In contrast to these three behaviors, negative emotional communication, lead thinking and other discussion of these three kinds of behavior occurs in the lower frequency but were significantly higher classroom behaviors. In the graduate's class, the frequency of positive emotions, ask questions and solve the problem was higher, but negotiation, produce, positive emotions, negative emotion, teacher organizing teaching and other discussions were significant classroom behavior in the graduate's class. In the face-to-face online synchronization class-room, undergraduate students and graduate students have a high level of attention and endurance. Except the discussion of cognitive dimensions, learners also appear the behaviors of expressing and exchanging emotions and discussing unrelated issues. Good emotional expression and communication provide a relaxed and friendly learning atmosphere for the formal teaching activities, which was conducive to the interaction of participants. The construction of high-quality deep knowledge can be realized in the learner's creation, reflection and solving problem, but the frequency of occurrence was less. Therefore, it is necessary for teachers to use teaching strategies to promote these behaviors.

In order to improve the quality of learners' online synchronization discussions, this study will provide some advice on teachers how to intervene in online synchronization discussions for undergraduate students and graduate students based on findings. There are four suggestions as follows.

First, for undergraduate students, teachers can encourage students to integrate their views and draw conclusions. They are more inclined to carry out the surface analysis so that they cannot think profoundly. Teachers can encourage students to integrate the various views that appear when students discuss to a certain degree. And then come to a conclusion or solution to make the whole discussion more complete. For graduate students, teachers can encourage students to reflect on their own point of
view. They prefer to express their own views, so teachers can guide students to reflect and summary on the point of view.

Second, teachers can encourage students to continue to ask questions after they have solved the problem, guide them to generate more knowledge building activities, and make positive responses and feedback to them.

Third, teachers can take more instructional strategy of teaching dimensions, such as organization of teaching, promoting discussion, guiding thinking and so on, to guide learners to share their views, argumentation, consultation and draw conclusions.

Finally, in formal learning activities, learners inevitably have some discussion that has nothing to do with the content of the teaching. At this point, the teacher can make appropriate reminders to avoid the whole discussion from deviating from the subject, thus improving the efficiency of learning and teaching.

This study combines the online discussion with the traditional classroom for blended learning and analyzes the behavior patterns of different learners. The research is very limited, so the researchers can do further research with various methods in different contexts.

References


Teachers' Concerns about Adopting Interactive Spherical Video-based Virtual Reality

Jie GENGa, Eric T. H. LUKb & Morris S. Y. JONGab *

a Department of Curriculum and Instruction, The Chinese University of Hong Kong, Hong Kong
b Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong
*lukeric@cuhk.edu.hk

Abstract: Fieldtrip-based learning is regarded as one of the most important pedagogic approaches in geography education. However, some valuable fieldtrip sites are too costly or and dangerous for school-age students to visit. With the technology of interactive spherical video-based virtual reality (ISVVR), low-cost virtual learning environments can be created to immerse students into near real-life contexts for conducting virtual fieldtrips. This study aimed to investigate geography teachers’ concerns about leveraging ISVVR in practice via an open-ended questionnaire-based survey. A natural language processing method, RAKE (Rapid Automatic Keyword Extraction), was adopted to analyze the collected data. Results have shown that the way to properly integrate ISVVR into the geography class is the peak concern. Directly acquiring proper ISVVR instruction resources can also influence teachers to adopt ISVVR. The technical support determines whether ISVVR can be successfully implemented in schools.

Keywords: Teacher Concerns, Interactive Spherical Video-based Virtual Reality (ISVVR), Geography Education

1. Introduction

The Road Map for 21st Century Geography Education: Geography Education Research proposed a research agenda over the next several decades. Within recommendations, 1) research about fieldtrip and its impact on learning geography knowledge, and 2) research about teacher preparation and the needs in geography teaching and learning were highlighted (Bednarz et al., 2013). However, some valuable fieldtrip sites are too costly or and dangerous for school-age students to visit. The virtual reality (VR) technology, as an advanced technology, allows users to explore immersive environments from any location and enables to construct and integrate different learning contexts (Pantelidis, 2009). The common concerns of the VR technology supported pedagogical approach would benefit for learning contexts where hazard and possibility of damage are involved. Spherical videos, also known as 360-degree videos, can record a view in every direction at the same time. This video-based virtual reality technology provides a new potential pedagogical approach. With the technology of interactive spherical video-based virtual reality (ISVVR), low-cost virtual learning environments can be created to immerse students into near real-life contexts for conducting virtual fieldtrips.

An innovation often requires not only gaining new skills but shifting one’s attitudes and beliefs. The effective implementation of an innovation is a highly personal developmental process. The Stages of Concern (SoC) Model, first was developed by Hall and Hord (2001), is a conceptual framework that describes, explains, and predicts probable participants’ behaviors through identifying their attitudes and beliefs toward an innovation. Further, Cheung and Yip (2004) refined the model with five dimensions that has been introduced in this study (See Table 1). This study was conducted in the context of teacher training programs. According to the SoC model, teachers’ concerns about leveraging ISVVR in practice were investigated, with the aim of better understanding on the necessary supports for the further program implementation.
Table 1: Operational definitions of the Refined SoC Model (Jong & Tsai, 2016, pp.333-334)

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Operational Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>The teacher is concerned about the worth and possibility of introducing the education innovation into his/her school.</td>
</tr>
<tr>
<td>Information</td>
<td>The teacher is uncertain about the pedagogical roles and tasks involved in the education innovation, and is concerned about the availability of corresponding support and resources.</td>
</tr>
<tr>
<td>Management</td>
<td>The teacher is concerned about the tasks and processes of implementing the education innovation; the issues related to efficiency, organizing, managing, scheduling, and time demands are utmost.</td>
</tr>
<tr>
<td>Consequence</td>
<td>The teacher is concerned about the impacts of the education innovation on student learning and his/her own professional development.</td>
</tr>
<tr>
<td>Refocusing</td>
<td>The teacher is concerned about the education innovation’s further development, exploring ways to improve the existing pedagogical effectiveness of the education innovation.</td>
</tr>
</tbody>
</table>

2. Research Design

2.1 Training of the ISVVR Technology for Geography Teachers

The subjects were from 60 secondary geography teachers who registered for the teacher development program. The training program was especially designed for leveraging ISVVR in geography classes. Two training events were organized. Each event consisted of 30 teachers. Two evening sessions for every training event were designed: 1) In Session 1, teachers first experienced spherical videos and discussed the importance of the fieldtrip in the Geography Curriculum. With the explanation of ISVVR relevant definitions and learning theories, a special topic of capturing 360-degree videos in a real environment was introduced. Then, participants (geography teachers) were divided into 5 groups. Under the training helpers’ guidance, they learnt on capturing 360-degree video fragments in practice. In the last part of Session 1, each group discussed the group assignment about designing a geography class with the ISVVR support. 2) In Session 2, the post-processing of ISVVR resources had been introduced first. Participants learnt the procedure of making ISVVR instruction resources. With uploading and downloading their ISVVR instruction resources (from YouTube to their mobile phones), low-cost virtual learning environments can be created by themselves. Google cardboard, as the last training part, can support future students using ISVVR instruction resources in their classes.

2.2 Survey Design

Given to the potential uncertainty when working with a smaller sample size (around 60 teachers), instead of conducting quantitative analysis, a survey with an open-ended question was designed: “The three most concerns when adopting ISVVR in the geography class”. The basic background information from responders was involved. Text-based data were collected.

2.3 Method for Text-based Data Analysis: Rapid Automatic Keyword Extraction (RAKE)

Natural Language Processing (NLP) is a crucial part of artificial intelligence for understanding complex languages. It involves a large variety of underlying tasks and machine learning models powering NLP applications. Keyword extraction is one of the most important NLP tasks when working with text (Shukla & Kakkar, 2016). By definition, keywords describe the main topics expressed in a document. Rapid Automatic Keyword Extraction (RAKE) is a well-known and widely used NLP technique. RAKE is an algorithm to automatically extract text-based keywords from documents. The algorithm follows strictly three steps of a typical keyword extraction algorithm, and has a good design structure (Berry & Kogan, 2010). In this study, we followed the original existing Python implementation of RAKE.
algorithm as described by Rose et al., (2010). The source code is released under the MIT License. The main steps are:

**Step 1. Candidate selection:** RAKE is based on the observation that keywords frequently contain multiple words but rarely contain standard punctuation or stop words (such as “and, the, of”, or other minimal lexical meaning). The input parameters comprise a list of stop words, a set of phrase delimiters, and a set of word delimiters. RAKE uses stop words and phrase delimiters to partition the document text into candidate keywords.

**Step 2. Properties calculation:** Co-occurrences of words within these candidate words are meaningful. Word associations are further measured. Enabling adaptive and fine-grained measurement of word co-occurrences is used to score candidate keywords.

**Step 3. Scoring and selecting keywords:** The RAKE score is a ratio \((\text{deg}(w) / \text{freq}(w))\) of the word degree \((\text{deg}(w))\) to the word frequency \((\text{freq}(w))\). The keyword identification is depending on the RAKE score. After candidate keywords are scored, the top \(T\)-scoring candidates are selected as keywords. Here, \(T\) is computed as one-third the numbers of words (Mihalcea and Tarau, 2004).

3. **Result**

### 3.1 Primary Dealing with Text-based Data

The survey was conducted at the end of each training event. Fifty-four attending teachers replied. The collected text-based data were a mix of English and Chinese. Since the implementation of RAKE is English-oriented, the data written in Chinese were translated into English before the analysis. The translation followed the rules of maximizing the accuracy: 1) it kept original sentence structure and the meaning of phrase using. For example, if a record was a phrase, it was translated as a phrase without any sentence extension. 2) Same words were translated by the consistent terms. As a result, 87 records about teacher concerns (from 54 teachers) were put into one text document (see Figure 1).

**Figure 1.** Records of teacher concerns about “Virtual Reality in Geography Class”

<table>
<thead>
<tr>
<th>Question: The three most concerns when adopting ISVVR in the geography class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responder 11:</td>
</tr>
<tr>
<td>1. After students turn on their mobile phone, they will play with their phones as they prefer;</td>
</tr>
<tr>
<td>2. Not enough technical support;</td>
</tr>
<tr>
<td>3. I am worried about those students may not listen to teacher instructions after turning on their mobile phone.</td>
</tr>
<tr>
<td>Responder 12:</td>
</tr>
<tr>
<td>1. Students' mobiles are not modern enough to use relevant VR app.</td>
</tr>
<tr>
<td>2. Very slow Wi-Fi connections that make the VR class resource difficult to be downloaded.</td>
</tr>
<tr>
<td>3. Students are doing other things when the time they should learn from VR resources;</td>
</tr>
</tbody>
</table>

### 3.2 Descriptive Analysis of Teachers’ Background

From the descriptive analysis of teachers’ background, within 54 teachers, 59% were female geography teachers and 31% were male geography teachers (9% were missing information). The age was well distributed: 11 teachers were 30 years old or below (20%); 18 teachers were with the age around 31-40 (33%); 19 teachers were around 41-50 years old (35%), and 6 teachers were 50 years old or above (11%). Furthermore, 59% of teachers did not have any experience of the VR technology; 31% of teachers had a little experience, and 9% of teachers were using the VR technology sometimes.
3.3 Keywords Extraction of Text-based Teacher Concerns

All extracted candidate keywords fulfilling with the T-scoring measurement were analyzed. Thirty-nine candidate keywords were generated by first running of RAKE. For refining the number of keywords, the second time running RAKE on these 39 candidates was conducted. The top-ranking keywords from two times running results were the same. Hence, all 21 keywords from the second time running results were selected. Since RAKE cannot recognize synonyms, 11 keywords were identified after manually combined with synonyms. In addition, through reading the original data again, keywords, 'transfer the knowledge', 'students' motivation', and 'learning effectiveness', cannot be recognized by RAKE, because they were only mentioned once and the phrase length was less than the average. Plus these 3 manually assigned keywords, as a result, 14 keywords were finally determined. The extracted keywords were grouped following the definitions of SoC model categorical concerns. Freq. (IR) recorded the number of the records that can be retrieved from original data with indexing keywords (see Table 2).

<table>
<thead>
<tr>
<th>SoC Model</th>
<th>Freq. (IR)</th>
<th>Keywords (Extracted by RAKE)</th>
<th>RAKE Scoring</th>
<th>Keywords (Manually assigned and combined with synonyms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>15</td>
<td>'VR class resource difficult' 14.14 'relevant VR app'</td>
<td>14.14</td>
<td>-- (false positives)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'relevant VR app' 8.94</td>
<td>8.94</td>
<td>'relevant VR app'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'software availability' 8.5</td>
<td>8.5</td>
<td>'modern mobile phones'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'modern mobile phones' 8.5</td>
<td>8.5</td>
<td>'modern mobile phones'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'smart phone availability' 8.5</td>
<td>8.5</td>
<td>'school support'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'school budget support' 7.5</td>
<td>7.5</td>
<td>'school support'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'VR device' 5.44</td>
<td>5.44</td>
<td>'VR device'</td>
</tr>
<tr>
<td>Information</td>
<td>5</td>
<td>'preparing VR instruction resources' 13.94</td>
<td>13.94</td>
<td>'preparing VR instruction resources'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'VR instruction resources' 9.94</td>
<td>9.94</td>
<td>'proper VR resources availability'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'proper VR resources availability' 13.69</td>
<td>13.69</td>
<td>'proper VR resources availability'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'teaching materials availability' 8.0</td>
<td>8.0</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'VR resources' 6.19</td>
<td>6.19</td>
<td>--</td>
</tr>
<tr>
<td>Management</td>
<td>10</td>
<td>'proper integrate VR instruction resources' 18.44</td>
<td>18.44</td>
<td>'proper integrate VR instruction resources'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'integrate VR resources' 10.19</td>
<td>10.19</td>
<td>'longer class hour needed'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'longer class hour needed' 14.7</td>
<td>14.7</td>
<td>'longer class hour needed'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'class order management' 9.2</td>
<td>9.2</td>
<td>'class order management'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'classroom order management' 9.0</td>
<td>9.0</td>
<td>'classroom order management'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'VR class preparation' 9.14</td>
<td>9.14</td>
<td>'VR class preparation'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'longer time preparation' 9.0</td>
<td>9.0</td>
<td>'longer time preparation'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'class preparation' 5.7</td>
<td>5.7</td>
<td>--</td>
</tr>
<tr>
<td>Consequence</td>
<td>4</td>
<td>'listening teacher instructions' 8.5</td>
<td>8.5</td>
<td>'listening teacher instructions'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
<td>--</td>
<td>'transfer the knowledge'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
<td>--</td>
<td>'students' motivation'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
<td>--</td>
<td>'learning effectiveness'</td>
</tr>
<tr>
<td>Refocusing</td>
<td>0</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

3.4 Validation of the Text-based Data Analysis

In the field of Machine Learning, NLP, and Information Retrieval (IR), the evaluation is a necessary procedure. To evaluate the performance of RAKE, the developers (Rose et al., 2010) tested RAKE against with experiments reported in Hulth (2003) and Mihalcea and Tarau (2004) by the same collection of technical abstracts. This benchmark evaluation clearly revealed that RAKE effectively extracted keywords in terms of precision, efficiency, and simplicity. Apart from the test for individual
documents, Rose et al. (2010) tested RAKE for a corpus of documents. Finally, the developers concluded that RAKE can be used in many applications where keywords can be leveraged.

In this study, precision, recall, and F-measure were applied as well. Out of 21 extracted keywords, 18 are true positives. There are 3 false positives (although they were extracted by RAKE, these keywords cannot be recognized as meaningful keywords.) in the set of extracted keywords, resulting in a precision of 85.71%. The 11 RAKE extracted keywords to the total of 14 identified keywords results in a recall of 78.57%. Equally weighting precision and recall generates an F-measure of 81.98%.

4. Discussion

According to the study of Cheung and Yip (2004) and Jong (2016), if teachers intensively concern of Evaluation, Information, and/or Management perspectives, no matter the innovation is widely recognized as a future tendency, teachers tend to reject to adopt the education innovation. If teachers concern more about Consequence and Refocusing perspectives, teachers are positively reflecting on the impacts of the innovation on students’ learning and on their own professional development. Once the peak categorical concern is identified, improvement on targeting their real needs of adopting the innovation should be formulated (Jong & Tsai, 2016).

4.1 Peak Teachers’ Concerns about “Virtual Reality in Geography Class”

Top 5 RAKE scoring keywords were selected to express the main content of a document (Medelyan, 2015). According to their belonging categories of the SoC model, both Information and Management are the peak categorical concerns from participating teachers (see Table 3).

Management is the top categorical concern of the SoC model. From the information retrieval, teachers are caring the most about the new geography class design with the ISVVR technology integration. The categorical concern of Information is the second major concern. From the information retrieval, the time to prepare ISVVR instruction resources is one of the major concerns. Meanwhile, directly acquiring proper VR instruction resources also influences teachers to adopt the ISVVR technology.

Table 3: Information retrieval by indexing keywords from the peak categorical concerns

<table>
<thead>
<tr>
<th>Categorical Concerns</th>
<th>Information Retrieval by Indexing Keyword</th>
</tr>
</thead>
</table>
| Management Freq. (IR): 3 | **Keyword:** ‘proper integrate VR instruction resources’ & ‘integrate VR resources’  
Responser 41: “when conducting the class design, we should proper integrate VR instruction resources. It is somehow difficult.”  
Responser 56: “How to integrate VR resources into the class?”  
**Keyword:** 'longer class hour needed'  
Responser 16: “longer class hour needed” |
| Information Freq. (IR): 5 | **Keyword:** 'preparing VR instruction resources' & 'VR instruction resources'  
Responser 41: “There is no IT technician to support the VR instruction resources. For doing it by myself, there is no time”.  
Responser 43: “Preparing VR instruction resources”;  
Responser 46: “Currently, there are not enough VR instruction resources. We need time to prepare.”  
**Keyword:** 'proper VR resources availability' & 'teaching materials availability'  
Responser 48: “Proper VR resources availability”;  
Responser 12: “Teaching materials availability” |
4.2 The Most Frequently Mentioned Categorical Concern: Evaluation

Fifteen out of 34 retrieved records by indexing extracted keywords are from the Evaluation category (see Table 4). “Whether the school is technically ready” is the most concern. From the perspective of school implementation, the technical support determines the successful implementation of ISVVR in schools, such as Wi-Fi, VR devices, relevant software, smartphone support, and budget support.

Table 4: Information retrieval by indexing keywords from the categorical concern of Evaluation

<table>
<thead>
<tr>
<th>Categorical Concerns</th>
<th>Information Retrieval by Indexing Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td><strong>Freq. (IR): 15</strong></td>
</tr>
</tbody>
</table>
| **Keyword:** 'relevant VR app' & 'software availability'; **Freq. (IR): 7**  
| Responder 40: “students' mobiles are not modern enough to use relevant VR app”  
| Responder 1, 2, 10, 14, 18 & 21: “VR hardware and software availability”. |
| **Keyword:** 'modern mobile phones' & 'smartphone availability'; **Freq. (IR): 3**  
| Responder 10: “Students' smartphone availability”  
| Responder 13: “WIFI, SMART PHONE availability”  
| Responder 56: “Not all students have modern mobile phones to use”. |
| **Keyword:** 'VR device'; **Freq. (IR): 2**  
| Responder 38: “VR device cost”  
| Responder 45: “VR device” |
| **Keyword:** 'school support'; **Freq. (IR): 3**  
| Responder 1 &2: “school support”  
| Responder 56: “School budget support”. |

4.3 SoC Model-based Categorical Concern: Consequence

In the categorical concern of Consequence, only the index keyword 'listening teacher instructions' was extracted by RAKE, the most concern is the students’ self-management (e.g. “students are not listening teacher instructions but playing their mobile phones”). Other three keywords, 'transfer the knowledge', 'students' motivation', and 'learning effectiveness', were manually assigned through reading the original data again. These keywords although rarely occur, they are important to highlight some teachers concerned about the learning effectiveness by leveraging ISVVR in practice.

5. Implications

5.1 Needs for the Proper Curriculum Design with the Advanced Technologies Integration

Advanced technologies and internet of things are changing the way we live and work, as well as the way of teaching and learning. In the early of 1998, Youngblut has mentioned that the Virtual reality (VR) technology would potentially become a major technological innovation and can offer significant support for education. Although the unique advantages of the VR technology in education are allowing students to perceive the near real-life learning contexts, the proper curriculum design with the advanced technology integration is still in challenging. Teachers still worried about the class preparation time and the availability of VR instruction resources once they would like to adopt this advanced technology. Even the instruction resources are ready, the proper integration of the advanced technology into the classroom has still influenced the teacher to adopt this education innovation.

Here, especially for the ISVVR in geography classes, a new teaching package with the proper ISVVR integration would be recommended. Geography teachers would be guided with an integral curriculum design. Pre-preparation of ISVVR instruction resources for some classic geography classes would be helpful. Teachers may first use the common ISVVR resources. After they adapt this pedagogical approach, they would like to explore the way to create their own ISVVR instruction resources.
5.2 A Conflict Between the Current IT Deployments in Schools and the Rapid Development of Advanced Technologies

The technical support determines the successful implementation of ISVVR in schools. The results revealed that there still has been a conflict between the current IT deployments in schools and the rapid development of advanced technologies. The technical support is the major constraint for the implementation of modern education innovations. The availability of the hardware and software, and the school budget support were the major concerns.

From this perspective, the increased education budget will be considered as one solution for promoting the pedagogical uses of future advanced technologies. Further, another tendency is able to foresee that with more and more free software and affordable hardware occupying the market, the instruments (both the hardware and the software) themselves should not be a serious challenge. Instead, the user experience and the UI design of the instruments, especially focusing on the pedagogical uses, will become a challenge for the technology R&D institutions once they want to expand their users in the education field. On the other hand, in near future, how to make a connection with the technology R&D institutions and the education institutions will determine the successful implementation of an education innovation.

5.3 Performance of the NLP Technique for the Text-based Data Mining

Back to the data analysis part, normally, a number of discrete information would be collected among the class observations, interviews, or some questionnaires. Text-based data is a common type of the data collected. The NLP technique enables to conduct the “big” and “discrete” text-based data mining. Instead of manually assigning or reading a huge number of the text-based data, the NLP technique would be helpful at the primary stage to generally picture these data.

In this study, it is also an attempt to use a NLP technique supporting the traditional qualitative analysis. RAKE as a typical NLP technique helped the keywords extraction from the text-based data. This study used a smaller sample size as a pilot test that the manually assigned keywords was possible to be conducted for the validation as well. With the keywords extracted from both RAKE and the manual method, the validation measurements have been conducted: a precision of 85.71%, a recall of 78.57%, and an F-measure of 81.98%, have shown a relatively good performance when dealing with the text-based data. One limitation of RAKE leads to the missing information from some rarely appeared but also important records. In addition, lacking synonym recognition causes additional manually assigned works required. Comparing with the time for manually organizing the text-based data, RAKE may help to deal with a big text-based data within a short time. For the perspectives of the primary checking and organizing the collected data, we are satisfied with the performance. The obtained results under the in-depth education theoretical framework, somehow, answered our research questions. As mentioned above, more NLP techniques will be explored to deal with the text-based data in near future. The NLP technique is possible to be a supplementation for the further qualitative analysis, especially under the conditions of the continuously generating of a big text-based data.

6. Conclusion

This study investigated geography teachers’ concerns about leveraging ISVVR in practice via direct observations in the training class and a survey with an open-ended question. Through the RAKE implementation and manually correction, 14 keywords were determined and grouped into the SoC categories. The results have shown that both Information and Management are peak categorical concerns. This means that teachers are still considering the feasibility of adopting ISVVR in practice. The way to proper integrate ISVVR into the geography class is the peak concern. Directly acquiring ISVVR teaching resources can strongly influence teachers to adopt ISVVR. The technical support determines the successful implementation of ISVVR in schools.
Acknowledgements

We would like to thank the participating teachers and our colleagues who helped the survey conduction.

References


How Designer Think About Designing an Augmented Reality App for the Study of Central Nervous System

Fadzil Saleh MOHAMAD ROFIE & Mas Nida MD. KHAMBARI*
Faculty of Educational Studies, University Putra Malaysia, Malaysia
*khamasnida@upm.edu.my

Abstract: The purpose of this qualitative case study is to explore the process of designing an augmented reality application for the study of the Central Nervous System. The researcher wants to gain an understanding of the process that take place during the entire of development process in developing the augmented reality application. Semi-structured interviews were conducted to seek the detailed process of developing the application and the designer’s interaction with the medical faculty members. The participant of the study is a designer who have relatively high experience in designing various software and developing games. Five themes emerged from the data analysis namely; (1) Planning, (2) Analysis, (3) Design, (4) Implementation, and (5) Maintenance.

Keywords: Augmented Reality, SDLC, Waterfall Model, Educational Software, Designer, Central Nervous System

1. Introduction

Mobile applications or commonly referred as apps are considered as one of the fastest growing trends in Information System Industry (Eddy, 2011). The number of effort to apply mobile technology in the learning process has rapidly increased over the years. The education system is also increasingly being supplemented by augmented reality technology to make learning more attractive and useful for learners. Using an augmented reality technology can offer an exclusive opportunity for educators to embed learning and supply three dimensional in a natural learning environment (El Sayed, Zayed, & Sharawy, 2011). The combination of these three concepts: augmented reality, education, and mobility will open great possibilities that can be utilized in numerous ways.

The rapid growth of mobile computing technologies alongside mobile software applications has made ubiquitous learning possible (Johnson, Adams, & Cummins, 2012). The major affordances of mobile computing technologies for learning include user mobility resulting from device portability, relatively stable computing power in small devices, and always on connectivity (Hsu & Ching, 2013). These affordances lead to the various potential for innovative uses of mobile technologies in education. The development of mobile apps has also raised interests among educators because it facilitates teaching and improves student’s motivation to learn (Johnson et al., 2012). Nowadays, Augmented Reality (AR) and Virtual Reality (VR) applications have an enormous popularity and significance with offered opportunities in concern with providing information to the sensory channels of individuals.

It is imperative to know what is in the mind of the designers while they are designing the application. As in many processes of building and developing computer software and information system, the waterfall model of SLDC has been used widely in industrial and engineering fields (Thayer & Yourdon, 1997). In relations, knowing whether the designer feels that they are following the steps in tandem with the waterfall model can illuminate some of the issues or a gray area that might be overlooked when creating the AR application. This is particularly the case for AR as this approach is still novel in the field of education. Besides that, there is a lack of research being done on interpersonal understanding between designers and the process of developing the application. This research is conducted in the hope of exploring this issue. The experience during the process of developing AR application can shed light on what is in the minds of the developer. Most development processes follow a very specific approach, and the Waterfall model is one of them. Knowing this will help understand
the development process of AR better and in return contribute to improving the application of AR in education.

2. Study Background

Recent developments in the field of technological advances have led to a renewed interest in transforming the educational processes from traditional methods to contemporary styles. Such advanced educational technology has improved the process of teaching and learning through the existence of hypermedia, mobile technology and the most recent is e-learning (Sriram, 2011). To satisfy the needs of educational technology, different adaptive methods are developed and practiced in the teaching and learning process. In addition, several types of research are conducted to analyze different educational software requirements and techniques used in the development software (Costa, Loureiro, & Reis, 2009; Riley, 2007). The use of educational technology is believed to have a profound effect in providing a student-centered learning environment. Similarly, Flemmer (2007) found that there are numerous areas in which the use of educational technology may benefit the students in the way of improving and satisfy the needs of students. Although there were various researcher studies conducted on the widespread use of educational technology, very few of them are focused on the full potentials of how educational software can contribute to the teaching and learning process. Against this background, it is necessary to examine in depth the relationship between the educational requirements and the software dynamics.

The rapidly emerging technological advancement influences the teachers’ or educators’ perspective in educational approach. The use of AR to revolutionize the teaching pedagogy in schools has become a worldwide trend (Chai, Wang, & Chiang, 2014). AR can be understood as an environment in which virtual and real elements appear to coexist. Similarly, in a widespread survey of Augmented Reality, AR is summarized as a field that lets the real world to be seen by the user, where it is superimposed or composited with virtual objects (Azuma, 1997). AR allows for a seamless combination of virtual content with the real world (Azuma, Billinghurst, & Klinker, 2011). The possibilities of integrating AR into education are endless as it can be applied to different types of learners and in a variety of fields ranging from language study to complex biological experiment such as medical simulation and high-risk experiment. An analysis by Hamilton and Oleneva (2011) explores several AR applications considered for teaching, including those developed for various industries such as media and entertainment, gaming, tourism and travel, social networking, marketing, and enhancement of daily life.

In the context of this study, the research is focused on the development of AR application for the Central Nervous System. At Universiti Putra Malaysia, medical students are usually required to attend both classes and laboratory sessions where they will be introduced to the 3D model of the human anatomy. The problem arises as the medical students have difficulty in visualizing 3D anatomy from the textbook. Most of the time in visualizing human anatomy, the students heavily rely on the lecturer’s creativity as well as how the lecturer delivers the learning process. In the previous research done by Shirazi and Behdazan (2013), the level of engagement between students in STEM education is gradually decreasing as they cannot understand the abstract concepts. Realizing this situation, an AR applications is developed to overcome such challenges and enhance the teaching and learning process. Through this application, it is hoped that the medical students will be able to understand the human anatomy in detail as they will be able to interact with the visual 3D human anatomy model. Thus, this study was conducted to understand the designer’s perspectives on how the process of developing an AR application takes place.

3. Research Questions

The purpose of this study is to explore the process of developing an AR application for the study of the Central Nervous System. Specifically, this project paper aims to answer the following research question, “How does the process of developing an augmented reality application for Central Nervous System take place?”
4. Methodology

Data Collection

Qualitative research design was employed for this study where data were collected through semi-structured interviews and document analysis. This method has been provided as it is believed to provide a better understanding of the process of designing an AR application and garner information on the designer’s expertise in the process of developing the application. This is supported by Merriam and Tisdell’s (2016) views which claimed qualitative research design is the most appropriate to use when the study involved understanding the process from an individual’s perspectives. Using a qualitative methodology, the researcher played the role as human instruments which are immediately responsive and adaptive are ideal means of collecting and analyzing data (Yin, 2014). Besides that, the researcher has also refer the participant as Ahmad; a pseudonym given to protect his identity (Patton, 2015).

In this study, the researcher has used semi-structured questions to draw out the responses from the participant. The participant was interviewed on the basis of an interview guide questions in which he was asked to describe the process of developing AR application. The interviews were conducted in three separate sessions over the cause of five weeks and each session lasted for about 40 to 60 minutes. Each interviews was audio recorded then transcribed to preserve the responses of the participants.

Besides interviewing, documents analysis was also employed. The researcher specifically used Ahmad’s personal documents to be part of documents that need to be analyzed. Besides that, the researcher has also analyzed the photography and media in the documents gathered as they are also considered as visual documents. This is in line with Pink (2013) which stated digital images are considered as part of contemporary reality due to its increasing presence in the world nowadays. The researcher has analyzed the designer’s photo album where most of the photos show the process of developing the AR application from beginning to the end. Through this technique, the researcher can observe the important of cultural values and story conveyed from that particular photo (Pink, 2013).

Selecting Research Participant

A purposive sampling technique is employed as it is as one of the effective ways in exploring the case specifically in the social domain context (Tongco, 2007). The researcher used this technique to ensure the selected participant has prior knowledge and expertise in the subject matter, which focuses on the development of AR application. In a similar context, Chein (1981) explains the purpose of purposive sampling is not to have an average opinion but rather the participants are carefully selected because of their unique experience and competence. The researcher knew the participant, Ahmad, through his supervisor. Ahmad was contacted and briefed about the purpose of the study. He then agreed to participate in the study. Only a single participant is used for this study as Ahmad is the sole designer involved in designing and developing the AR application for the Central Nervous System. From the initial plan of the project, the designer is responsible for the development phase until the software is delivered to the stakeholder, which is the Faculty of Medical and Health Sciences, Universiti Putra Malaysia. Ahmad can be described as a fully dedicated, inspiring person and has a vast knowledge in the development of software.

5. Data Analysis

In this study, the systematic grounded theory analysis has been adopted to analyze the data collected. The data have been analyzed by using the open coding method to enable the researcher look for patterns and trends (Yin, 2014). This technique was also adopted to ensure that the data are not left behind and important information are not overlooked. Before starting data analysis, the researcher has carefully listened to the interviews and then transcribed them by using the verbatim technique. Then, the transcribed interviews were read several times to enable the researcher to familiarize with the data. The researcher has also marked the meaning units he considered appropriate for the purpose of reading, discussing, and reaching consensus. Other than that, the researcher has adopted several different techniques in analyzing data including writing memos on what the researcher learns, try out ideas and themes on participants, and by exploring new literature while in the interview session (Anderson, 2010).
The researcher started conducting the intensive data analysis to ensure that it is sufficient to answer the outlined research questions earlier.

Adhered to Yin’s (2014) suggestion that a qualitative research requires a careful data management of data collection, the researcher has used colorings notes to set out thoughts, hunches, and speculations to data collection. Besides that, the researcher also provides a log book to act as personal inventory which has records of field notes, interview data, documents, and memos. The inventory then is organized and labeled accordingly to enable easy access to the researcher. To interpret data collected from the interviews, the researcher has categorized the findings into several themes. The formation of themes are based on different process based on the waterfall model. This is in line with Creswell (2013) who suggested development of themes are required and recommended in the data analysis procedure. By using member checking technique, the researcher is able to construct reliable and valid themes to be used in this study (Merriam & Tisdell, 2016). Through this technique, the researcher has asked the participants to read and view the transcripts and documents. The researcher has asked the participants to verify that all transcripts did not have any additions other than what he had stated in the interviews. This is similar with Patton (2014) who stated that member checking involves the process of taking data and interpretations back to the people from whom the data were derived to confirm the credibility of the information.

6. Findings

This section will discuss the findings related to the compelling research question, “How does the process of developing an augmented reality for the study of Central Nervous System take place?” Five main themes have emerged from the data analysis, namely (i) planning, (ii) analysis, (iii) design, (iv) implementation, and (v) maintenance.

6.1 Planning Phase

In the planning phase, the medical faculties together with the designer needed to identify the requirements to determine the objectives of this project. A simple application has been developed as a preliminary experiment based on the requirement given by the medical faculty. To gain a better understanding of this requirement, this preliminary output has been demonstrated to medical lecturers from the Faculty of Medical and Health, UPM along with the Computer Scientist lecturer from Faculty of Computer Science and Information Technology, UPM. The goal of this phase is to develop an initial understanding of the new system or software. The designer stated that in the planning phase, the medical faculty would explain the problem. From the discussion, it was decided that the application should display a human brain in a 3D model. Alongside the 3D model are some information and a set of questionnaires. This is evident in the following excerpt:

“At this phase, the stated problem was the medical students need a model of the human brain when they want to study about the Central Nervous System. Because of the limited availability of this model, they only can use the model in the laboratory only. They cannot learn outside the laboratory as the model is not there. The purpose of this apps is to provide them access and the model of the brain when they want to learn anytime.”

[Interview 1]

During the planning phase, the designer highlights two important things. First, the importance of user requirements and second, the dynamic of user requirements to the learner’s perspective as well as educator’s perspective. This is where the designer works together with the medical faculty in addressing any discrepancies between the educational needs and software that need to be developed. In the second interview, the designer mentioned this issue:

“In the planning phase, for me, user requirements are very important. Because the requirements show what the problem is, what learner’s needs to be addressed and what the educators want. This is why I have several meetings with the medical faculty and science computer faculty to clarify the requirements. They also need my expertise to
explain the available technology and how to integrate it with the proposed solution.”

[Interview 1]

According to Boehm (1970), the first phase of Waterfall Model is called “System Feasibility” where the planning phase focus on the development proposal that consists of the formation of the project timeline, initial project estimation plan, and communication plan. In this case, the designer stated that the importance of project timeline and the communication plan that involve the medical faculty as well:

“In the beginning, I create the Gantt chart where there is a specific period for every phase. For example, week 1 and week 2, I will only focus on the planning phase. Week 3 until week 13 I will focus on implementation phase.”

[Interview 1]

“To avoid any misunderstanding during the planning phase, I create a Whatsapp group where I can communicate directly with the medical faculty. If there is greater problem occur, I will personally meet-up with them.”

[Interview 2]

To proceed to the next phase, the designer explains the importance of clear understanding about the proposed solution. The designer felt that it is vital to not waste time by solving the wrong problem. Throughout the several meetings, in the beginning, the designer must have a clear understanding of the given project and work with the medical faculty in creating the new system.

6.2 Analysis Phase

The goal of this phase is requirements determination where the designer required to understand the requirements of the proposed apps and to develop the proposed apps as required by the medical faculty. The designer also mentioned the need of drafting the level of detail for the system to be functioning.

“Medical faculty provides me the user requirement, where it has the objective of the apps, why it the apps need to be developed and the required function of the apps. They also provide a draft interface on how the apps should be functioning”

[Interview 1]

In this phase, the designer also mentions the importance of analyzing the current applications on the market to get a statement of what the application must do and the characteristics of the proposed apps must have. Figure 1 exhibits some of the existing AR application in the market. The designer commented as:

“In this phase, I was required to explore the various type of article and journal, and I also do some applications review on the current market. This is to give me more information on how the apps should be, the design interface and the content.”

[Interview 1]

In addition to the analysis requirements, the designer highlighted the process of defining the project scope such as user scope, content scope, and system scope:

“Based on my previous experience, when I need to design a new system, I normally divided it into three scopes; First user scope, then contents scope and system scope. For user scope, I need to analyze the user requirement given by the medical faculty. Then in content scope, I will look on the storyboard provided, and the system scope is where I will prepare a design for the system and later present to the stakeholder.”

[Interview 2]
6.3 Design Phase

The design phase looks at how the architecture of the system will be built and how the system will function in particular emphasis on software, hardware, user requirement, and application interface. The designer explains the purpose of this phase which required the development of a blueprint that will satisfy the requirement as given by the stakeholders.

“In the design phase, I will create a storyboard which is the same as a blueprint. The storyboard I produce is based on the requirement given by the medical faculty. From the complex requirement, I make it simple through the storyboard. I also make a draft interface for the design application, around 20 pics of interface and then I present it to the stakeholders.”

[Interview 1]

He also added that the blueprint would identify the input and output needed and the overall process of how the application should behave. The designer mentioned in the following comment:

“The storyboard will have a draft interface on how the application will look like, the specific button and how it will interact with the content. If the stakeholders satisfied with the blueprint, then I don’t have to add anything.”

[Interview 3]

6.4 Implementation Phase

This phase is focused on the physical development of the system as set out in the blueprint during the design phase. Normally in the implementation phase, the project required a development team that consists of programmers, testers, interface designers, and other experts. However, in this case, the designer will solely act as a programmer, interface designer, and tester specialist. The designer also highlights six key activities in this stage which are slightly different to normal activities as the implementation stage should be. this is explained in the following excerpt:

“In my previous experience where I have design games, there are only 3 to 4 activities during the implementation stage. I thought there is no different when designing an AR app. However, through the development process, I notice that there are different activities that I need to complete in order to finish this phase.”

[Interview 3]

The implementation phase starts with digitization process of the real model is turned into a 3D model. Figure 2 shows the process of digitization and the result of the model. The designer explained the process of digitization as follow:
“First step is modeling which I called as digitization process. Using the real model of the brain that supplied by the medical faculty, I turn it into a 3D model. At first, I used a software called Kinect; later I found it not compatible with others software. That is when I used software called Autodesk Remake. The digitization process is where I need to take a picture of the brain model from every angle, from top to bottom.”

[Interview 3]

Figure 2. The process of digitization.

Coding is where the physical design and development of specifications are turned into working computer code. The designer also described the process of video recording as shown in figure 3. The designer explains the development process of coding and video recording activities:

“In the coding phase, I will create an interaction by using the draft interface such as picture, buttons, symbol and the digitized model to link to an action. For example, if I click this button then it will bring to the next interface or users can interact with the 3D model. I use Unity software which allows me to create a programming environment to develop the AR apps.”

[Interview 3]

“Based on the requirements given by the medical faculty, I was given a task to develop a video specifically for this app. I need to set up a schedule with the medical lecturers and do a video recording of lecturer given a lecture on Central Nervous System.”

[Interview 3]

Figure 3. The video used in CNS app was recorded at the Museum of Anatomy in Universiti Putra Malaysia

Next step in the implementation phase is testing where the designer needs to confirm whether the system meets the requirements as set out by the stakeholders. In this phase of testing, the designer will perform a validation process with the medical experts. All of them have the experience of teaching Anatomy of Central Nervous System subject for more than five years. The designer said overall suggestions and comments had assisted him greatly in the process of making this application a success. The designer also described the medical expert as the first level user during the testing process:

“In the testing phase, I need to do a validation process to confirm the function of the apps to tell with the requirements of the stakeholders. The first step in the testing phase is when I did validation with the medical experts. They will try the apps and then give
feedback on the function and content used.”

[Interview 3]

Also, the designer also mentioned the need for re-testing process which includes the medical students as a second level user.

“After the testing process with the medical experts, I need to make some improvements. Later after one week of improvements, I do re-testing with 12 students from medical faculty in order to get feedback from their perspective.”

[Interview 3]

After the re-testing process, the next key activity is documentation that consists of feedback from users and medical experts. In this phase, the designer will compile the feedback and improvements given by the end-user to finalize the apps. The designer also highlighted the implementation phase as the longest compared to other phases:

“For me, this phase is the longest and most difficult compared to other phases. Because I took a longer time to find the right technique for digitization process. I should have completed it in week ten however due to the difficulty in digitization process; I completed it in week 13 as I only discover the right technique.”

[Interview 1]

“When I stuck at implementation phase, I feel not very confident enough whether I can deliver this project or not. This is where I got the problem in digitizing the model and turn it into a 3D model. Looking back at my timeline, after week ten then I only got a solution to my problem, and I got so happy. I can say that this was the most difficult process.”

[Interview 2]

6.5 Maintenance Phase

The final phase of the Waterfall model is maintenance. As the new system is implemented, it will often need ongoing maintenance and support. The designer stated that the maintenance phase in vital to ensure the designed apps is operating at optimum levels of performance.

“The main activities in this phase includes getting feedback from the end users and analyze the feedback in order to make changes. From the previous phase, where I have conducted a pilot test with the medical students, the result of the test is being analyzed to ensure the application is operating within the requirements as specified.”

[Interview 2]

However, the designer mentions that there is no maintenance phase for this project. This is due to the requirements stated by the stakeholders. The designer commented as follow:

“There is no maintenance phase for this project. I was given a project where I only need to develop only a prototype. The maintenance phase is where you have the official product and later need ongoing maintenance.

[Interview 1]
7. Conclusion

In understanding designer’s thinking when designing an AR application for the study of the Central Nervous System, this study had found that the design process went through different phases of the Waterfall model. This study revealed that software methodology play significant roles to serve as a development framework to the designer. This is consistent with McMurtrey (2013) as he found the use of methodology can give a great impact on the success of a software development. Through this study, the researchers also found the Waterfall model adopted by the designer shows that different phases and approaches affect the process and quality of the AR application. It is clearly proven that each software methodology has significantly different characteristic that makes them suitable for different type of projects.

Other than that, the designer in the interview also emphasizes that designer’s understanding towards the project’s requirements facilitate the selection of appropriate SDLC for a particular project. Similarly, Mahanti, Neogi, and Bhattacherjee (2012) also stated the aspect of understanding user requirements, project complexity and project team are associate with the SDLC’s selection. The finding is consistent with past studies conducted by Chakraborty, Arefin, Baowaly, & Bahar, 2012; Mahanti et al., 2012; Sriram, 2011) which highlights user requirements as the most important factor in selecting an SDLC model. In this study, the researchers found that the Waterfall model chosen by designer was chosen based on several factors. Apart from its clear framework, the designers claimed that the use of this model is more effective in the time factor when compared to other model. In conclusion, this study found that the result of clear requirements in planning phase provided by the medical faculty give a huge impact on the development process and its successful product delivery.

References


Technologies (EDULEARN09) (pp. 5816-5825). Barcelona: International Association of Technology, Education and Development (IATED).


A Comparison of Different Types of Learning Activities in a Mobile Python Tutor

Geela Venise Firmalo FABIC*, Antonija MITROVIC & Kourosh NESHATIAN

Computer Science and Software Engineering, University of Canterbury, New Zealand
*geela.fabic@pg.canterbury.ac.nz

Abstract: Programming (i.e. coding) is becoming one of the skills expected for successful careers in the knowledge economy¹, and is being taught at all levels, including primary and secondary schools. Programming skills are difficult to acquire, as the student needs to learn the specific programming language and many related concepts to write good programs. We present PyKinetic, a mobile tutor for Python that serves as a complement to traditional courses. The overall goal of our project is to design learning activities that maximize learning. In this paper, we present several types of learning activities designed for PyKinetic. The first version of the tutor implemented Parsons problems with incomplete lines, which support code-understanding and code-writing skills. The second version of PyKinetic included various types of activities aimed at code-tracing and code-writing skills. The results of two studies we conducted show that Parsons problems are beneficial for novices, while advanced students benefitted more from learning activities which required them to identify and fix incorrect lines of code.

Keywords: Mobile Python tutor, Parsons problems, self-explanation, code writing, code tracing, code understanding

1. Introduction

Smartphones are mainly used for communication, but are also widely used for leisure and entertainment as they provide ubiquitous access to most services. Smartphones may also prove to be an effective learning platform. In New Zealand, 72% adults have access to or own a smartphone and/or a laptop (Research New Zealand, 2015). A trend also emerged that 59% of people with multiple devices prefer using smartphones. Furthermore, the same study revealed that millennials (aged 18-34) are the highest smartphone users in New Zealand at 91%. Most novice programmers nowadays are millennials who expect fast-paced interactions (Oblinger and Oblinger, 2005). Educators therefore need to investigate alternative avenues of learning that could be more effective and appealing to millennial novice programmers. As trends continue, smartphones may prove to be an effective platform to learn programming, as it provide opportunities to learn while “on the go”.

Python is widely used nowadays as a programming language in introductory programming courses (Guo, 2013). We present PyKinetic (Fabric, Mitrovic and Neshatian, 2016a; 2016b), a mobile Python tutor, developed using Android SDK to teach Python 3.x programming. The tutor is a complement to traditional lecture and lab-based introductory programming courses. Being a mobile tutor, we hope that it would appeal better to a new generation of students, compared to desktop or Web-based educational tools. Traditional code writing exercises may be difficult on a small-screened device such as a smartphone, as the keyboard usually obstructs half of the smartphone screen. For that reason, we have designed learning activities for PyKinetic that require only tap and long-click interactions.

The overall aim of our project is to design activities that will maximize learning (Fabric, Mitrovic and Neshatian, 2017). In this paper, we present a set of learning activities that focus on code-understanding, code-tracing and code-writing skills. The first version of PyKinetic (PyKinetic_IncLOCs) contains only Parsons problems (Parsons and Haden, 2006), which are exercises requiring the student to re-arrange a given set of randomized Lines Of Code (LOCs) to produce the expected outcome. This version of the tutor contains Parsons problems with incomplete LOCs, in which

¹ https://obamawhitehouse.archives.gov/blog/2016/01/30/computer-science-all; www.codefuture.org
the student needs to re-arrange the LOCs and complete the missing elements. In this way, PyKinetic_IncLOC supports both code understanding and code writing skills. To further support learner engagement, we have also added self-explanation prompts. Self-explanation (SE) was defined as generating inferences and interpretations from principles not taught explicitly by the material, and has been shown to increase learning (Chi et al., 1989; Wylie and Chi, 2014). We have conducted an experimental study investigating the learning effectiveness of PyKinetic_IncLOCs. Our hypothesis was that PyKinetic_IncLOCs would be successful in supporting learning (S1_H1). We also hypothesize that self-explanation prompts would result in additional learning benefit (S2_H2).

The second version (PyKinetic_DbgOut) contains debugging and output prediction problems, designed to support acquisition of debugging and code tracing skills. We conducted a study using this version, hypothesizing that the combination of activities in this version of the tutor will also be beneficial for learning Python programming (S2_H1).

The goal of this paper is to compare the learning effectiveness of the two versions of the tutor, focusing on types of students who benefitted from those learning activities. The paper is structured as follows. In the following section, we present related work on Parsons problems. Section 3 presents the study conducted using PyKinetic_IncLOCs. Section 4 presents learning activities in PyKinetic_DbgOut, as well as the results of Study 2. Section 5 compares the results from both studies and discusses what kind of students benefitted from the learning activities in those two studies.

2. Parsons Problems

Parsons problems were originally proposed as a fun way for learners of Turbo Pascal to improve their syntactic skills (Parsons and Haden, 2006). These puzzle-like activities are suitable for novices, as they already contain syntactically correct code that only needs to be put in the right order. Variations of Parsons problems include extra LOCs (distractors), or incomplete LOCs which require the learner to provide missing elements. The latter is implemented in the version of PyKinetic presented in this paper.

There were other variants of Parsons problems considered (Denny et al., 2008). Two variants did not contain any distractors, and the only difference between the two is that one of them includes scaffolding such as curly braces and indentation (since this was used in the context of Java), while the second variant does not provide any. Two other variants (available with and without scaffolding), provided paired options for each LOC which were given in randomized order, with the paired options given right next to each other. The last variant still contains pairs of options for each LOC. However, every LOC including the paired options, were given in a randomized order. It was not specified whether the last variant was presented with or without scaffolding, but this variant ended up being discarded as it was perceived to be unreasonably difficult. For example, because of the paired options, seven lines of code for the puzzle becomes 14. Having twice the amount of LOCs in a completely randomized order may be overwhelming for students.

More research is encouraged to pinpoint accurately which skills benefit most by Parsons Problems. Some believe Parsons problems to be simpler than code tracing (Lopez et al., 2008) while some find that based on its complexity, Parsons problems lie between code tracing and code writing (Lister et al., 2010). Code writing requires higher order skills, while code tracing falls into lower categories in Bloom’s taxonomy (Thompson et al., 2008). Moreover, a weak correlation was found (Denny et al., 2008) between scores on Parsons problems with code tracing questions, and a moderate positive correlation with code writing. Denny et al. (2008) suggested that Parsons problems were similar to code writing. A recent study (Morrison et al., 2016) found that Parsons problems pose a lower cognitive load compared to code writing, which may be due to the correct syntax given in Parsons problems. However, this may not be always true, as Parsons problems may require higher cognitive load depending on the type, complexity, and interface used (on paper or on a device). Moreover, there are opinions that the position of Parsons problems in the hierarchy of programming skills can vary, depending on their type (with or without distractors) and complexity (Ihantola and Karavirta, 2011). More factors that could influence learning are scaffolding and feedback provided.
3. Study 1

3.1 Parsons Problems in PyKinetic_IncLOCs

We have chosen Parsons Problems as the first activity for PyKinetic as they are simple activities for novice programmers, also suitable for a smartphone interface. Parsons Problems in PyKinetic are completed by dragging and dropping single LOCs in the correct order. Karavirta, Helminen, and Ihantola (2012) also perceived Parsons problems to be suitable for mobile devices and have developed MobileParsons for Android and iOS. Solving Parsons problems on a smartphone means learners are not required to use the keyboard to type their answers.

PyKinetic_IncLOCs contains Parsons problems with incomplete LOCs (Figure 1, left). The student fills an incomplete line by tapping to see alternatives and selecting one of them, like the implementation by Ihantola, Helminen, and Karavirta (2013). We believe that solving Parsons problems with incomplete LOCs would enhance learners’ code writing skills more than Parsons problems with or without distractors. Furthermore, a recent study (Harms, Chen and Kelleher, 2016) found that Parsons problems with distractors decrease learning efficiency of middle-school children aged 10-15.

We conducted a study with PyKinetic_IncLOCs, using 15 problems covering six Python topics: string manipulation, conditional statements, while loops, for loops, lists, and tuples. All problems required LOCs to be rearranged to match the given problem description and expected output. The student had to complete the current problem to proceed to the next problem.

The first two problems were used as practice, to familiarize participants with the mode of interaction supported by the system. The remaining 13 problems had between 3 and 16 LOCs, with a maximum of three incomplete LOCs per problem. The problems were given in a fixed order of increasing difficulty, with initial problems having a smaller number of LOCs, and a smaller number of incomplete LOCs. Furthermore, the first half of the problems focused on a single topic each, while the other problems covered at least two topics. The initial seven problems were code snippets, while latter problems were more complex: each consisted of a function with function calls. PyKinetic_IncLOCs recorded information about all actions performed by participants in system logs.

![Figure 1. Example of a problem in PyKinetic_IncLOCs (left) and SE prompt (right)](image)

To further improve learning, we introduced menu-based self-explanation (SE) prompts (Wylie and Chi, 2014). Figure 1 (left) provides an example Parsons problem, with one incomplete line (the
highlighted line). The code must be rearranged to match the given expected output, which is *PyKinetic says hello!* After completing that line, the student is given an SE prompt (right screenshot in Figure 1), which can only be attempted once to help prevent trial and error. Each SE prompt starts with the completed line *(message = “says hello”)* in Figure 1, right), and asks the student to select correct (or incorrect) statements related to that line. The screenshot in Figure 1 (right) illustrates a situation in which the student selected only the first option, but that selection was wrong. There can be more than one correct option for each SE prompt. The tutor provides feedback for every wrong option (Figure 1, right), and highlights correct options in green and incorrect options in red.

There were 22 SE prompts in total, 14 of which were conceptual questions and 8 were procedural questions. All incomplete LOCs must be completed before the solution to the Parsons problem can be submitted. Learners receive two types of feedback for submitting a Parsons problem, either: “Correct! Great job!” for a complete solution (including completed LOCs) or “Check the order of your solution.”

### 3.2 Experimental Design

There were two conditions in the study: the only difference between the versions of PyKinetic_IncLOCs presented to the control and experimental group was the additional SE prompts in the experimental condition. Our first hypothesis was that all participants, irrespective of the group, would improve their Python skills by solving Parsons problems (S1_H1). Secondly, SE prompts would help experimental group participants learn more than control group (S1_H2).

We recruited 47 volunteers enrolled in COSC121, an introductory programming course at the University of Canterbury (UC), and 13 volunteers from a local high school (HS). The high school participants were taking a Year 13 course on Digital Technology. We also recruited 23 volunteers enrolled in an introductory computing course at the Ateneo de Manila University (ADMU). There were 83 participants in total, randomly assigned into two groups: experimental group (with SE prompts) and control group (without SE prompts). The study was approved by the school principal, and Human Ethics committees of UC and ADMU.

Each participant participated in a group session that lasted for 1.5-2 hours. There were one up to 13 participants per session. At the start of each session, the participants were introduced to the study and provided informed consent. Afterwards, a 15-minute pre-test was administered on paper, and the participants were instructed on how to download and install the tutor. After working with the tutor, the participants received a 15-minute post-test, also administered on paper. Lastly, instructions were given on uninstalling and deleting the application. Some participants used their own Android smartphones, while we provided phones to other participants.

The UC participants have learned previously all topics covered in PyKinetic_IncLOCs in their course. Although the ADMU participants have not learned about tuples, they were advised to attempt all 15 problems (including practice questions) within the time limit of an hour. We have later confirmed that the ADMU participants were not disadvantaged on the pre-test, as there were no statistically significant differences between the results of UC and ADMU participants on the pre-test question about tuples. High school participants were instructed to attempt 13 problems only, because they have not learned about tuples, and we had very limited time constraints with the high school (the sessions were conducted during strict time-scheduled periods of 50 minutes). Due to the same reason, high school participants received pre-tests on a different day of the same week, and the rest of the study was conducted on another day in 50 minutes.

The pre/post-test had eight questions each: six conceptual questions (6 marks) and two procedural questions (2 marks). The conceptual questions were multiple-choice or True/False questions. One procedural question asked the participants to predict the code output, and the other one was a Parsons problem. Questions with multiple correct answers were marked depending on the options selected. Partial marks were given for selecting correct options, and for not selecting wrong options. Partial marks were deducted for selecting wrong options. This was done to avoid discrepancy for participants who seemed to be guessing answers by selecting all options. Parsons problems from the pre/post-test were marked based on the number of LOCs written in the correct order combined with expert knowledge. Parsons problems in the tutor itself were not marked. Only the SE prompts were marked using the same marking scheme used for multiple-choice questions in pre/post-test.
3.3 Findings

There was no significant difference on the pre-test scores between the three populations of students (UC, HS, ADMU), thus showing that they had similar pre-existing knowledge. Table 1 reports the pre/post-test scores of the experimental and control groups on all questions, and on conceptual/procedural questions separately. We used the paired non-parametric Wilcoxon Signed Ranks test to verify hypothesis S1_H1. Both experimental and control groups improved scores from pre- to post-test overall (the Improvement row), as well as on conceptual questions (the Improvement Conceptual row). For procedural questions, there was no significant improvement. These results show that there is enough evidence to accept our first hypothesis S1_H1, which was that PyKinetic_IncLOCs would be effective in supporting learning.

Table 1. Statistics from Study 1 (at p < .05: * denotes significant; ns denotes not significant)

<table>
<thead>
<tr>
<th></th>
<th>Experimental (36)</th>
<th>Control (40)</th>
<th>U, p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time/problem in tutor (min)</td>
<td>4 (1.57)</td>
<td>3.18 (1.13)</td>
<td>U = 502, p = .023*</td>
</tr>
<tr>
<td>Pre-test %</td>
<td>64.75 (18.52)</td>
<td>66.01 (12.34)</td>
<td>ns</td>
</tr>
<tr>
<td>Post-test %</td>
<td>75.86 (16.15)</td>
<td>70.56 (14.37)</td>
<td></td>
</tr>
<tr>
<td>Improvement</td>
<td>z = -3.315, p = .001</td>
<td>z = -2.45, p = .014</td>
<td></td>
</tr>
<tr>
<td>Cohen’s d</td>
<td>d = .64</td>
<td>d = .34</td>
<td></td>
</tr>
<tr>
<td>Normalized Gain %</td>
<td>14.94 (77.79)</td>
<td>4.82 (63.71)</td>
<td>U = 530, p = .048*</td>
</tr>
<tr>
<td>Pre-test Conceptual %</td>
<td>62.40 (19.05)</td>
<td>63.41 (14.31)</td>
<td>ns</td>
</tr>
<tr>
<td>Post-test Conceptual%</td>
<td>75.71 (16.91)</td>
<td>69.19 (16.71)</td>
<td></td>
</tr>
<tr>
<td>Improvement Conceptual</td>
<td>z = -3.221, p = .001</td>
<td>z = -2.37, p = .018</td>
<td></td>
</tr>
<tr>
<td>Pre-test Procedural %</td>
<td>71.82 (26.98)</td>
<td>74.42 (19.48)</td>
<td>ns</td>
</tr>
<tr>
<td>Post-test Procedural %</td>
<td>76.17 (21.42)</td>
<td>74.58 (19.95)</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 1 also reports the results of the Mann Whitney U test for checking significant differences between the two groups. There was no difference on the pre-test scores, but the experimental group performed significantly better on the post-test (U = 529.5, p < .05). There was also a significant difference on the normalized gain (U = 530, p < .05). Both groups had a positive Cohen’s d effect size, but the effect size was higher for the experimental group (experimental: d = .64; control: d = .34). These results support our second hypothesis S1_H2, which was that participants who self-explained would learn more than the control group. Participants from the experimental group spent significantly more time per problem in comparison to the control group, which was expected, as they needed to answer SE prompts (p < .05, U = 502).

Table 2. Effect sizes for novices/advanced students

<table>
<thead>
<tr>
<th>Group</th>
<th>Ability</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Novices (20)</td>
<td>51.76 (12.78)</td>
<td>71.34 (17.48)</td>
<td>d = 1.28</td>
</tr>
<tr>
<td></td>
<td>Advanced (16)</td>
<td>80.99 (9.30)</td>
<td>81.51 (12.68)</td>
<td>d = .05</td>
</tr>
<tr>
<td>Control</td>
<td>Novices (18)</td>
<td>55.23 (7.99)</td>
<td>65.58 (15.06)</td>
<td>d = .86</td>
</tr>
<tr>
<td></td>
<td>Advanced (22)</td>
<td>74.84 (7.05)</td>
<td>74.64 (12.70)</td>
<td>d = -.02</td>
</tr>
</tbody>
</table>

We performed a post-hoc split of participants based on the median of the pre-test scores (66.47%). Participants who scored less than the median were labelled as novices, while the rest were considered as advanced participants. As presented in Table 4, there was a big difference between the effect sizes for novices and advanced participants in both groups. The effect sizes were very small for advanced participants in each group, while there were substantially higher effect sizes for novice participants. The novices from the experimental group obtained a higher effect size than novices in the control group (Table 2). Moreover, novices in the experimental group had a significantly higher normalized gain than novices in the control group (U=120, p = .08).
4. Study 2

4.1 Learning Activities in PyKinetic_DbgOut

The problems in PyKinetic_DbgOut consisted of the problem description, code (containing 0-3 incorrect LOCs), and 1-3 questions. There were five types of questions (Table 3): three types of debugging questions, and two types of output-prediction questions.

Table 3: Five Types of Debugging and Output Prediction Questions in PyKinetic_DbgOut

<table>
<thead>
<tr>
<th>Type of Question</th>
<th>Additional Information Given</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbg_Read</td>
<td>Test cases with actual output</td>
<td>Is the code correct? (Yes or No)</td>
</tr>
<tr>
<td>Dbg_Ident</td>
<td>Test cases with actual output</td>
<td>Identify ( n ) erroneous LOCs (( n ) is given)</td>
</tr>
<tr>
<td>Dbg_Fix</td>
<td>Test cases with expected output</td>
<td>Fix erroneous LOCs (by tapping through given choices)</td>
</tr>
<tr>
<td>Out_Act</td>
<td>Test cases</td>
<td>Select actual output of the code</td>
</tr>
<tr>
<td>Out_Exp</td>
<td>Test cases</td>
<td>Select expected output of the code</td>
</tr>
</tbody>
</table>

In \( \text{Dbg\_Read} \) questions, the learner is given some test cases with the actual output; the learner’s task is to specify whether the given code is correct or not. The second type of debugging questions (\( \text{Dbg\_Ident} \)) provides similar information to the learner, but requires the learner to identify one or more incorrect LOCs. An example is shown in Figure 2 (left screenshot), where the student needs to identify two incorrect lines (the lines the student selected are highlighted in blue). The third type of debugging questions is \( \text{Dbg\_Fix} \), which starts with requiring the student to identify incorrect lines (\( \text{Dbg\_Ident} \)), and then to fix them (Figure 2, right). To fix incorrect LOCs, the student needs to select the correct option from given choices. In the screenshot shown in Figure 2 (right), the student has completed the line highlighted in green, and is working on the other line (highlighted in orange).

Each output-prediction question contains 1-3 test cases. In the first type (\( \text{Out\_Act} \)), the student needs to specify the actual output of the given code for each given test case (Figure 2, middle). For example, if the code is erroneous the actual output may be none with an error displayed (Figure 2,
middle, last option). On the contrary, in Out_Exp questions, the student specifies the expected code output matching the problem description.

PyKinetic_DbgOut had 21 problems provided in a fixed order. There were seven levels of complexity, each containing 2-4 problems (Table 4). Problems on levels 1-3 cover conditionals, string formatting, tuples and lists; these problems consist of 4-8 LOCs (excluding function definition, comments, and test cases), and only one question. For example, problem one is a code-reading problem, containing only a Dbg_Read question. The complete code, problem description, and test cases with function calls are given; the task is to identify if the code is correct or not.

Every problem in levels 4-7 each contained 2-3 questions, and covered same topics plus for loops, while loops, and importing a module. Each problem on these levels started by requiring the student to identify incorrect LOCs (Dbg_Ident). After that, levels four and five had output prediction questions next: identifying the actual output (Out_Act) for level four, and identifying the expected output (Out_Exp) for level five. Level six targets code writing skills, by requiring the student to fix erroneous LOCs (Dbg_Fix) in the second question. Lastly, level 7 contains three types of questions in each problem: identifying erroneous LOCs (Dbg_Ident), identifying actual output (Out_Act) and fixing erroneous LOCs (Dbg_Fix). The problem illustrated in Figure 2 belongs to level 7. It is important to note that the ordering of the problems is not solely reliant on the number of LOCs and topics involved in the problem. In some cases, the code and/or the problem itself may be more logically complex than others even though it had fewer lines and topics.

<table>
<thead>
<tr>
<th>Level</th>
<th>Problems</th>
<th>Additional Information Given</th>
<th>Topics Covered</th>
<th>Number of LOCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dbg_Read (2 problems)</td>
<td>Test cases with actual output</td>
<td>Conditionals</td>
<td>4-6</td>
</tr>
<tr>
<td>2</td>
<td>Dbg_Ident (4 problems)</td>
<td>Test cases with actual output</td>
<td>String Formatting and Conditionals</td>
<td>4-8</td>
</tr>
<tr>
<td>3</td>
<td>Out_Act (4 problems)</td>
<td>Test cases</td>
<td>String Formatting, Conditionals, List, Tuples</td>
<td>4-8</td>
</tr>
<tr>
<td>4</td>
<td>Dbg_Ident -&gt; Out_Act (2 problems)</td>
<td>Test cases</td>
<td>String formatting, Conditionals, List, Tuples, For loops</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Dbg_Ident -&gt; Out_Exp (2 problems)</td>
<td>Test cases</td>
<td>String formatting, Conditionals, Lists, For loops</td>
<td>8-9</td>
</tr>
<tr>
<td>6</td>
<td>Dbg_Ident -&gt; Dbg_Fix (3 problems)</td>
<td>Test cases with expected output</td>
<td>String formatting, Conditionals, Lists, For/While loops, Importing a module</td>
<td>9-11</td>
</tr>
<tr>
<td>7</td>
<td>Dbg_Ident -&gt; Out_Act -&gt; Dbg_Fix (4 problems)</td>
<td>Test cases</td>
<td>Nested While loops, Conditionals, Lists, Tuples and String Formatting</td>
<td>11-16</td>
</tr>
</tbody>
</table>

4.2 Experimental Design

We conducted a study with PyKinetic_DbgOut, with recruited 37 participants enrolled in COSC121, which was the same course where we recruited most of our participants for Study 1 (Section 3.2). We have eliminated data about two participants as they have not finished the study. The sessions were two hours long, with 1-9 participants per session. The participants provided informed consent, followed by an 18-minute pre-test, which included questions on demographics and programming background. We then gave brief instructions on using the tutor, and provided Android smartphones with the tutor already installed. Participants interacted with the tutor for roughly an hour. Lastly, participants were given an 18-minute post-test either when time had run out or when they had finished all problems. The post-test included open-ended questions for comments and suggestions about the tutor. The study was approved by the Human Ethics Committee of the University of Canterbury.
The topics covered in the study have previously been covered in COSC121 lectures. The pre- and post-tests had six questions each and were administered on paper. The tests contained same types of questions from Table 3 (worth one mark each), and additionally a code-writing question (worth 5 marks). The participants were not used to doing any programming exercises on paper, because all lab quizzes and assessment in COSC121 are completed using computers. Therefore, code syntax on their pre/post-test were not strictly penalized. There were no multiple-choice questions in the pre/post-tests. The code-writing question provided the problem description, test cases with expected output, function definition statement and the docstring. The code-writing question from both tests had an ideal solution of 5 LOCs (without any comments), which was the reason for a maximum of 5 marks on this question. The participants did not receive scores for the problems completed in the tutor.

4.3 Findings

The results from Study 2 are presented in Table 5. There were no significant differences on any reported measures. The problems the participants were solving in the tutor are of different nature to the problems they were used to in the course, where they were asked to write code. For that reason, we investigated whether there is a difference between the participants based on their code-writing skills. Before Study 2, the participants were assessed in a COSC121 lab test, which consisted of 20 code-writing questions. The median score on the lab test was 79%. We therefore divided the participants post-hoc into two groups based on the lab test median: we refer to the 16 participants who scored less than 79% as novices, and to the 19 participants who scored 79% or higher as the advanced students.

Table 5. Pre/post-test scores (%)

<table>
<thead>
<tr>
<th>Question</th>
<th>Pre-test%</th>
<th>Post-test%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total score</td>
<td>68.55 (23.28)</td>
<td>72.88 (24.67)</td>
</tr>
<tr>
<td>Dbg_Read</td>
<td>77.14 (42.6)</td>
<td>74.29 (44.34)</td>
</tr>
<tr>
<td>Dbg_Ident</td>
<td>88.57 (32.28)</td>
<td>80 (40.58)</td>
</tr>
<tr>
<td>Dbg_Fix</td>
<td>57.14 (46.8)</td>
<td>60 (44.64)</td>
</tr>
<tr>
<td>Out_Act</td>
<td>93.57 (23.75)</td>
<td>90.71 (26.49)</td>
</tr>
<tr>
<td>Out_Exp</td>
<td>89.05 (23.89)</td>
<td>78.1 (30.99)</td>
</tr>
<tr>
<td>Code Writing</td>
<td>56 (40.09)</td>
<td>69.14 (36.17)</td>
</tr>
</tbody>
</table>

Table 6 presents the results for novices and advanced students. Advanced students were expected to perform better than novices, as they started with a higher level of existing knowledge. Indeed, that was the case: advanced students outperformed novices by completing more problems (U = 35, p < .05) and by getting higher pre/post-test scores. Although the overall pre-test score was significantly different for the two subgroups of students, this was not the case with the score on the code-writing question alone, where there was no significant difference between novices and advanced students. Furthermore, only advanced students improved their score on the question for fixing erroneous code (z = -2.51, p = .012) and on the code writing question (z = -2.07, p = .039); the novices have not improved on those questions. Lastly, it seemed that output prediction questions were unfavorable for the learning of advanced students, as there was a significant difference between novices and advanced on the pre-test but no significance in the post-test.

We investigated further on whether there was a correlation between the average time spent per completed problem and the normalized gains. The normalized gain on all questions for novices was moderately positively correlated to the time spent per problem (r = 0.52, p < .05), but the correlation was not significant for advanced students. Contrary to that, there was a strong positive correlation between the normalized gain on only code-fixing and code-writing questions, and time spent per problem for advanced students (r = 0.7, p < .001), and no significant correlation for novices.
5. Discussion and Conclusions

We presented two versions of PyKinetic, and the findings from the two studies. In Study 1, all participants interacted with Parsons problems with incomplete LOCs, but the experimental group participants additionally had to answer SE prompts. The results from Study 1 supported our hypotheses: that PyKinetic_IncLOCs was successful in supporting learning (S1_H1), and that the SE prompts provide additional learning benefits (S2_H2). Both groups improved their scores on conceptual questions from pre- to post-test; a potential explanation for this improvement may be that Parsons problems contribute to the acquisition of conceptual knowledge. Furthermore, SE prompts have been shown in previous studies to contribute to conceptual knowledge (Najar, Mitrovic and McLaren, 2016).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Novices (16)</th>
<th>Advanced (19)</th>
<th>U, p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed problems</td>
<td>19.63 (1.54)</td>
<td>20.53 (1.17)</td>
<td>U= 35, p=.037*</td>
</tr>
<tr>
<td>Time/problem (min)</td>
<td>2.67 (.80)</td>
<td>2.82 (.44)</td>
<td>ns</td>
</tr>
<tr>
<td>Pre-test (%)</td>
<td>58.39 (21.09)</td>
<td>77.11 (22)</td>
<td>U= 76, p=.011*</td>
</tr>
<tr>
<td>Post-test (%)</td>
<td>57.92 (28.3)</td>
<td>85.48 (10.75)</td>
<td>U= 63.5, p=.003*</td>
</tr>
<tr>
<td>Improvement</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Normalized gain</td>
<td>-.03 (.7)</td>
<td>.13 (.74)</td>
<td>ns</td>
</tr>
<tr>
<td>Pre-test Dbg Fix</td>
<td>34.38 (46.44)</td>
<td>76.32 (38.62)</td>
<td>U= 77, p=.012*</td>
</tr>
<tr>
<td>Post-test Dbg Fix</td>
<td>34.38 (42.7)</td>
<td>81.58 (34.2)</td>
<td>U= 62, p=.002*</td>
</tr>
<tr>
<td>Improvement Dbg Fix</td>
<td>ns</td>
<td>z= -2.51, p=.012</td>
<td></td>
</tr>
<tr>
<td>Pre-test Code Writing</td>
<td>45 (37.59)</td>
<td>65.26 (40.74)</td>
<td>ns</td>
</tr>
<tr>
<td>Post-test Code Writing</td>
<td>46.25 (41.13)</td>
<td>88.42 (14.25)</td>
<td>U= 76.5, p=.011*</td>
</tr>
<tr>
<td>Improvement Code Writing</td>
<td>ns</td>
<td>z= -2.07, p=.039</td>
<td></td>
</tr>
<tr>
<td>Pre-test Output Questions</td>
<td>84.11 (19.62)</td>
<td>97.37 (7.88)</td>
<td>U= 84.5, p=.024*</td>
</tr>
<tr>
<td>Post-test Output Questions</td>
<td>84.9 (17.86)</td>
<td>83.99 (21.17)</td>
<td>ns</td>
</tr>
<tr>
<td>Improvement</td>
<td>ns</td>
<td>z= -2.29, p=.022</td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, Parsons problems with incomplete LOCs proved to be effective for novice learners, but showed no effect for advanced learners. This outcome is consistent with Morrison et al. (2016), who found Parsons problems posed a lower cognitive load compared to code writing. Furthermore, based on our own experience in teaching Python, lower performing students usually have difficulties in writing their own code. Parsons problems provide sufficient scaffolding as the complete code is given with correct syntax, which only requires re-arranging. It is possible that advanced learners did not benefit from using PyKinetic_IncLOCs because they already had mental models of solutions. Hence, it is probably easier for advanced learners to write their own code based on their mental model. We have observed this with experts, when investigating strategies used in solving Parsons problems with distractors (Fabic, Mitrovic and Neshatian, 2016b).

Our hypothesis (S2_H1) for the study conducted with PyKinetic_DbgOut was that a combination of debugging and output prediction problems would also be effective for learning. However, our results revealed no significant improvement between pre- and post-test for all participants. This outcome might be due to the small number of participants in the study. Delving deeper, we found that, contrary to Study 1, PyKinetic_DbgOut proved to be more beneficial to advanced students, and showed no effect for novices. Code-writing and code-fixing exercises revealed to be more suitable for advanced students. Furthermore, a strong positive correlation was found for advanced students between the normalized gain on only debugging and code-writing questions and time spent per problem, but no significant correlation for novices. The correlations suggest that advanced students improve their code-writing skills more as they spend longer time on each problem, but minimal effect for novices. This was possibly because novices require more support.

The findings from the two studies would enable us to develop an adaptive version of PyKinetic. We plan to add a student model, which would be initialized based on the student’s result on the pre-test. The student model would be updated with every activity the student performed, and will enable the tutor to select learning activities for the student tailored per his/her student model.
Acknowledgements

We thank Dr. Ma. Mercedes Rodrigo and her team from Ateneo de Manila University for collaborating with us on Study 1, and Mr. Patrick Baker and Middleton Grange School for allowing us to conduct our study with their students.

References


A Case study of Young Children’s Use of iPad for Digital Storytelling for a Study of Self

Leigh DISNEY* & Gretchen GENG
School of Education, Charles Darwin University
*leigh.disney@cdu.edu.au

Abstract: Young children gain an understanding of self via play. Visual representations of self via art in early childhood are often associated with children’s drawing or painting; the current study reports on digital story telling, rarely used by children. Reasons vary, including the lack of user-friendly technologies and difficult analysis of young children’s self-learning process and results. This study was innovative by involving a case study of a 7-year-old child’s play with iPad and whether and how this child could reflect on self and present results by using visual artistic presentation other than traditional painting or drawings. This 4 week study found that the participating child could use apps, such as iPad app camera, to produce a digital story presenting her study of self. Although this study has its limitations, it does provide a new area for early childhood educators to explore and study further.

Keywords: early childhood education, visual artistic presentation, digital story telling, digital technology

1. Introduction

An important component of holistic development for young children is the development of an understanding of self (Berk, 2013). By developing an understanding of self, it allows children to understand how they fit within the world and how they can make sense of new information (Berk, 2013). One way in which children make sense of self is through artistic representations (Dockett & Perry, 2009). Traditionally, drawings are one of the most primitive forms of children expression and communication (Mitchell & Ziegler, 2007). Kress (1997) and Steel (1999) agree that drawings are regarded as effective means for children to explore their understandings, and the process of drawings and the accompanying narrative are integral parts of the meaning-making process (Cox, 2005; Einarsdottir, Dockett & Perry, 2009; Wright, 2007). Stanczak (2007) states that the meaning of the images represents most importantly in the way that children interpret those images. Cox (2005) further states that drawing is a constructive way of thinking-in-action. However, children have to be asked to explain their drawings to avoid adult interpretation; and young children’s views and experiences can be assessed by paying attention to their narratives and interpretations as they draw as well (Clark, 2005; Dockett & Perry, 2005; Punch, 2002; Veale, 2005). Therefore, it is not surprise that although drawings have been used as a method to interpret and report on children’s understanding of self, it is never ‘truly’ representative of children’s sense of self, as all the drawings have to be interpreted by adults.

With the advent of technological innovations such gestural interface devices, it is impacting the way children can represent visual arts. Visual arts are considered particularly important to children’s understanding of self, because they can represent a mirror to facilitate self-reflection, and force critical, highly meaningful and pleasurable consideration of the social and cultural dimensions of personal experience (Glass, 2011; Weber & Mitchell, 2004). Visual artistic expressions that have been used to interpret and report on self-study include (a) performance, (b) photography, (c) video documentary, and (d) multimedia representations. Although all these expressions considered as important methods in self-study, it has rarely been used effectively by young children learning. The reasons behind could include the above involved technologies are too difficult for young children to use, and it was not easy to represent children’s sense of self, particularly in early childhood settings.

The emergence of gesturally based tablet computers is changing the landscape for the integration of young children and technology. By offering a child-friendly touch screen-technology
learning environment, children’s play with tablets has become a topic that is of considerable interest for early childhood education audiences. Segal (2011), for instance, suggests that new technologies allow for new opportunities to include touch and physical movement, which can benefit learning in contrast to less direct methods such as somewhat passive interactions of a mouse and keyboard. Research by Chan and Black (2006) found that immediate sensorimotor feedback received through direct manipulation of animation through the hands allowed for increased learning as compared to passive animation conditions. Interactive gestures allow for this form of learning hence a greater likelihood of meaningful learning experiences (Saffer, 2009). Moreover, children observe adults such as their teachers and parents using digital technologies throughout their daily life (Donahoo, 2014). Children receive various kinds of information such as images, audio and textual information from digital technologies (Edwards, 2013). Therefore, the contexts in which children view and use digital technologies reflect on how they constructing knowledge about their world (DEEWR, 2009). Thereby children now have access to and are utilising emergent technologies, such as Apple’s iPad, as a means of recording and understanding their environments as a way to comprehend their worlds and consequently themselves.

One of the ways young children can use digital technology and gain an understanding of self is via digital storytelling. Mantei and Kervin (2010) describe digital storytelling ‘as an art form’ with Karagiannidou (2017, p.27) describing that digital storytelling allows learners to incorporate text, images, photographs, audio and video files, as well as other artefacts, music and narration’. Children are utilising a number of different visual artistic forms in order to reflect a sense of the environments in which the live, learn and grow. Furthermore ‘the affordances of digital technologies can make this form of expression and learning more interactive, immersive and even personal, one that can be created and, most importantly, shared outside a particular classroom or setting’ (Karagiannidou, 2017, p.27). Children develop a sense of personalization with their electronic devices which allows them to investigate their personal experiences via the use of touch screen video recording devices. Children are then able to record and present a representation of themselves that is told by them and requires no interpretation by adults. To date there is very limited research that highlights children’s independent use of emergent digital technology to act as storytelling device to make sense of self. The current study investigates a young child’s play with her tablet device in order to make sense of the important features of her home life.

2. Case Study

This study used a case study to demonstrate and provide useful references and reflective comments, because it was important to go into details describing children’s self-study via use of visual artistic presentation thoroughly, rather than researching or surveying into different opinions of whether or not this technology can be used, this paper accordingly includes a video documentary produced by the participant (7 year old girl), as a better idea which would be of high interests to readers. Case studies ‘focus on only one individual or one thing…which enables a very close examination and scrutiny’ (Salkind, 2009, p. 213). By turning to the discussion of this digital story, we cannot only provide some theoretical background about visual-informed research, but also the context of their application to children’s understanding of self.

Case studies allow for an in-depth analysis of a participant in a naturalistic setting allowing for a detailed reflection of occurring phenomena (Kervin, Vialle, Herrington & Okely, 2006). Case studies are most effective and rigorous when ‘utilising multiple data collection sources’ (Kervin, Vialle, Herrington & Okely, 2006, p. 70). Within this study naturalistic observations, interview and artefacts were examined to evaluate the participants engagement and motivations during the research process.

The participant of this study was a 7-year-old girl. She was normally developed, including cognitively, and socially. She was attending a local public school and receives mainstream education in classrooms. She had been using iPad since 3 years old, firstly at home and then at school. By the age of 5 she had access to her own iPad at home, with apps purchased and monitored by her parents and has been using and playing age appropriately developed learning apps, such as art, numeracy, literacy and science apps developed for early childhood age children (birth-8) or junior primary age children (5-9 years).

The participating child was provided an iPad herself to use and play for a period of 4 weeks with full access to the app “camera” on the iPad (which she had used previously) as well as a range of art
and story making Apps such as Drawing Pad and Toca Life: School. The apps downloaded on the iPad were all educational apps listed as recommended apps in Apple store. The prices were ranging from free ($0) to $5.99. She did not have access to account to purchase any new apps or produce any in-app purchase. The participating child was given the following instruction: can you make a story about yourself using the iPad? At the end of each week she was asked the questions: have you made a story about yourself yet? The participating child had the free access to the iPad during the 4 week period. She also had access to adult support, such as parents or teachers, in understanding the use of gestural interfaces of iPad.

3. Findings

This section presents the findings from the naturalistic observations, weekly questions, artefact analysis and post artefact interview. The observations were divided into weekly periods. The interview was conducted at the end of Week 4. The study was considered completed when participant developed a visual artistic presentation of self via use of the iPad.

3.1 Observations and Interview from Week 1

The participating child was very interested in all the apps provided for her and undertook free play with the apps, occasionally seeking help from adult support in regards to technical elements of gameplay, such as altering brush colours in the arts based app. The child would generally use the iPad for durations of 20-30 minute blocks with sustained attention given to gameplay.

This finding is consistent with the literature about young children's engagement while using computers; and young children can use iPad to play and learn.

At the end of week 1 after being asked: have you made a story about yourself yet?, the child responded “No, not really, I’m just having fun”.

3.2 Observations and Interview from Week 2

The child continued playing predominantly with the drawing app, which allowed her to record her final product. However the child was deleting each image after creation, including artistic representations of self, family and friends.

This finding is consistent with literature that indicates children can use art as a manner to reflect their image of self. However, at no stage did the child show any of the images to the parents in a manner of storytelling to illustrate self.

At the end of week 2 after being asked: have you made a story about yourself yet?, the child explained that she had drawn some pictures of herself but it wasn’t a story, rather “just some pictures I drew, I really liked the one of Dunlop [the cat]”.

3.3 Observations and Interview from Week 3

The child started to use the iPad camera to record images of herself and anything around her. These were predominantly still images to begin with. As the week progressed the child showed a very strong interest in herself by video recording her own facial expressions, laugh and speech. The child also used the camera to record things or happening around her. This is an very important step of young children to use iPad to start producing her own work based upon her own self-studying or understanding of surroundings. However, although she recorded videos which were important to herself, as an art presentation during her own studying, the videos or photos the participating child produced were either portrait of herself or simply recording with no self-illustration or much understand of the meaning of her own comprehension of self.

At the end of week 3 after being asked: have you made a story about yourself yet?, the child indicated that she had made “lots of movies”, but that ultimately she hadn’t finished yet.
3.4 Observations and Interview from Week 4

By week 4 the child had full technical grasp of how to use the camera and produced 23 short videos. During the fourth week the child started to record herself as a narrator in the videos. For example, she held the devices to record her surroundings with her own understanding. At the end of week 4 after being asked: *have you made a story about yourself yet?*, the child indicated that she was indeed ready, however that “*I need Mummy’s help to turn it into a proper movie*”.

Vygotsky’s zone of proximal development suggests that there are a range of tasks that might be too difficult for the child alone, but with the help of a more abled peer they will be able to achieve the task (Berk, 2013). In this case the child understood that she was capable of making the separate videos, however to turn them into a video she was going to need help. Subsequently the child along with the Mother used Movie Maker to turn the spate videos into a movie. The child selected the videos, text and music for the movie and the Mother assisted in the technical elements of movie making. Therefore finalising the child’s “movie” and consequently creating a digital story.

3.5 Artefact Analysis

From the 23 videos the child took, she selected three videos she called "home friends"; the participating child recorded her cat, her dog and her sister. In producing these short videos, she was starting to use descriptive narration such as "*this is my cat, her name is Dunlop, and she is a girl* [...] (see Figure 1)", to introduce important element of her life. She also gave detailed descriptions of her “friends’” characteristics, such as "*this is my sister, she is a silly baby, naughty* [...] (see Figure 2)" or "*[...] my dog is a grandpa (he is old) dog* (See Figure 3)". Although the quality of the videos was not great or professional, the participating child managed to use the video as an artistic way to present components of her home life that represent her understanding of self and articulate a digital story.

![Figure 1](image1.jpg) “This is my cat, her name is Dunlop, and she is a girl…”

![Figure 2](image2.jpg) “This is my sister, she is a silly baby…”
3.6 Final Interview

The participating child was interviewed after the study. The child emphasised the personal elements of digital storytelling (her cat had recently died, hence a motivation to record her videos) importance of her confidence of using the app “camera” and the reasons why she started to use her explanation while recording.

Interviewer: “Can you tell me why you started taking so many videos”?
Child: “Because Mango (family cat) had died and I wanted to take videos of my pets. I did my sister because she is funny!”

Interviewer: “Do you enjoy yourself while recording?”
Child smiled and nodded her head.
Child: “I did, I enjoyed walking around and taking videos of all those things”

Interviewer: “Why did you add the narration in the videos”?
Child: “I did “This was made by blaa, blaa” because I had seen this done in my class by other kids. I talked to let people know what I was doing”

Interviewer: “Does the video tell a story about yourself”?
Child: “Sort of, I think it tells a story about my company people and pets”.

Interviewer: “If you were to make another video, is there anything else you would do or do differently”?
Child: “Maybe record some bits if my house, mum and dad and I would add more words to tell people what I was doing.”

4. Discussion and Conclusion

With emergence of technologies, more and more user-friendly devices are made available to young children (Ebbeck & Waniganayake, 2010). This study agrees that this transformation could bring dramatic changes to those traditional methods of understanding young children's learning, in particular their understanding of self. Play is a pedagogy in early childhood education to provide good examples for young children to learn. Allowing children to engage with user-friendly tools allows them to present their own learning in meaningful and reflective ways. This study shows that as well as traditional artistic expressions such as drawings, young children (such as the 7 year old child in this study) can use emergent technological methods, such as digital story telling as visual artistic representation of self.

As this technology is slowly emerging, it should be noted that this study is only a start and should attract more detailed research involving theory and practices to benefit children in their learning and early childhood educators’ pedagogy.

This research has its limitations. One of the weaknesses of the case study maybe its interpretative nature. The researchers as the sole data analysers, with his or her subjective perspective,
may interfere with objective of results. Moreover, the research is only one reflection of practice and therefore has limited generalisability; yet still demonstrates the underlying foundation that young children can not only use emergent digital technology, but can represent images of self without the need for adult interpretation. Therefore, more research will be conducted in the future to explore in this new area and bring possible innovative pedagogical methods to allow children to present personaliised version of self within and outside of early childhood education settings.

References

Online Responses towards the Impact of Hand Held Devices on Children’s Social and Emotional Development

Gretchen GENG* & Leigh DISNEY
School of Education, Charles Darwin University
*gretchen.geng@cdu.edu.au

Abstract: This paper used responses from an online discussion forum to investigate people’s opinions towards the use of hand held devices for young children and their impact on children’s social and emotional development. Critical discourse analysis (CDA) was used as analysis methods in three-dimension framework via micro, meso and macro multi-level interpretations. Although online response showed different opinions towards the impact of hand held devices, all the 125 online responses agreed that young children should not overuse the devices, and they should have both technology time as well as non-technology time to meet development needs. Educators and parents, who are interested in the areas of whether or not young children should be exposed to and use hand held devices, will find this paper useful.

Keywords: digital technology, hand held devices, online responses, young children, social and emotional development

1. Introduction

Within the field of early childhood a recent change that has influenced the manner in which children play and learn has been the advent of emergent digital technologies. Within the new digital age, parents and educators have access to and are allowing children to be exposed to a wider variety of technological devices, such as hand held devices, to enhance and build on children’s experiences and develop their own pedagogical practices (Colker, 2011). Hand held devices, such as tables, iPads and mobile phones refer to those portable devices which are used for communication and entertainment purpose such as gaming, media and internet access (Balakrishnan et al., 2016). There are a perceived number of benefits of the technological devices for children, but not all educators and parents are welcoming of the integration of digital technology and there is debate over the place of digital technology in the lives of young children (Blackwell et al., 2013; Daugherty et al., 2014; Zaranis et al., 2013). Furthermore, there is an undercurrent that technological devices, particularly those hand held devices are drastically altering the landscape of childhood to its detriment (Plowman, McPake & Stephen, 2010).

The study of child development is integral to understanding the developmental capacities and potential of young children. For ease of discussion, development is often broken into categories, these being: social and emotional, physical and cognitive domains of development (Dodge, Colker & Heroman, 2002). Each an intertwining domain allows for a broad understanding of how children learn, develop and grow (Deewr, 2009). This paper is focused on children’s social and emotional development, which is a broad domain of development that covers how children feel about themselves and others (Dowling, 2014). It captures children’s play, work and attachment experiences (Parke & Gauvain, 2009). Allen and Marotz (2010, p. 41) explain that children’s ‘gender roles, temperament, independence, morality, trust, acceptance of rules, and social expectations’ are important components of social and emotional development. Longitudinal research shows correlation between children’s social and emotional developmental characteristics (e.g. temperament, self-control, and resilience) and long-term social and emotional benefits into young adulthood and beyond (Rudasill et al., 2010; Schoppe-Sullivan et al., 2009). Early childhood is seen as a time for children to develop the social and emotional skills needed for transition to school (Berk, 2013). The importance of social and emotional development on long-term positive learning outcomes for children is well documented (Brown, Winsor & Blake, 2012).
There is research that has identified the potential positive effects of technology on children’s social and emotional development within the field of education for children (Lee, 2015; NAECY & Fred Rogers Centre, 2012; Zaranis et al., 2013). Dowling (2014, p.33) states that ‘although mobile phones and screen technology are now more available to young children a lack of personal contact is certainly not evident in early years settings… Children chatter as they work and play’. In the study by Yelland, Gilbert and Turner (2014) with kindergarteners, the use of iPads in a play-based setting highlighted improved self-regulation and persistence whilst engaged with the touch screen technology. They also detailed that children who had difficulty sustaining concentration normally had no problems whilst engaged with the iPad.

Conversely, there is an undercurrent of believe that hand held devices adversely effects pre-schoolers’ social and emotional development (Armstrong, Donohue & Highfield, 2015; Brown et al., 2012; Tokmak, 2013). Opponents of the use of digital technology within childhood point to a number of factors for non-use. The social environment with embedment of technology is not achieved easily. Digital technologies are feared to be addictive and socially isolating, minimising time spent with families and peers, inhibiting language development resulting in poor communication skills that are needed for school and beyond, fostering negative social and emotional outcomes, for example, tantrum, fear and anxious behaviour, aggressive and disruptive behaviour, as well as taking time away from children’s imagination and engagement with books (Bower, 2013; Brown et al., 2012; Finegan & Austin, 2002; McCarrick & Li, 2007; Plowman & McPake, 2013; Plowman & Stephen, 2003; Rosenfeld, 2015; Schewartz, 2011; Struppert, Guo & Waniganayake, 2010; Yelland, 2011). Overuse of digital technologies could cause developmental issues in young children (Woods & Miano, 2012).

Although many studies have been conducted using digital technologies effectively in educational environments so that their negative potential is averted and their positive potential is promoted, it was never showed whether the stakeholders, such as parents, educators, medical practitioners as well as technicians had knowledge about the impact of hand held devices on children’s social and emotional development and its related behavioural outcomes. This study therefore, is to analyse open online responses to reflect the current attitudes and beliefs from the key stakeholders towards the impact of hand held devices.

2. Methods

The Internet is an “alternative medium for voice amplification by those often without access to or control of mainstream media” (Tamatea, 2010, p. 984). The online responses from the bulletin board or chat groups may not represent the general public opinions towards the use of hand-held devices in children’s learning. However, the Internet-sourced data provides a map upon which various opinions and attitudes can be located and compared from different perspectives and concerns. There are a lot of research papers written based upon the analysis of online data, such as Tamatea (2008; 2010) and Geng and Disney (2014).

An author posted an online discussion platform about the use of handheld devices should be banned for children. One hundred and twenty-five readers who had access to Internet and the online discussion contributed to the discussion of agreeing or disagreeing with the author’s opinion and the possible background of their responses. The online data was entered by different people around the world from February to May, 2014 and highlighted the nature of the present discussion in relation to the children’s use of digital technology that is available online.

Using similar analysing methods – critical discourse analysis (CDA) in Geng and Disney (2014), this report analysed different online responses towards children’s use of hand-held devices. CDA is based upon both linguistic theory and social theory. Linguist theory elaborated by Ainsworth and Hardy (2004), Henderson (2005), and Wodak (2001) is supporting CDA through a three-dimensional framework – micro, meso and macro-level interpretations. Moreover, theorists such as Habermans (1990) already addressed that critical social theory helps CDA approach to examine ideologies and power relations involved in discourse. The micro level interpretation was around the topic of whether digital technology should be banned for children. The meso level interpretation was mainly focused upon the benefits and harm from the use or overuse of the technology. The macro level interpretation of the online data however concentrated on the strategies which shall or should be used in
children’s use of digital technology. Although the data reported what could be considered opinions of the people without much research or literature support or development, the data is the opinions of the people who speak for or on behalf of children’s parents, educators or medical practitioners, and who care to contribute on the discussion platform.

3. Online Responses

The 125 online responses were mainly focused upon an argument concern, which was whether the hand held devices could cause issues, such as delayed development, obesity, sleep deprivation and mental illness in young children. Among the responses, 48.8% of the 125 total responses (61 responses) were related with impact of hand held devices on children’s social and emotional development.

Out of the 125 online response, 22 responses (17.6%) strongly agreed that the hand-held devices impacted negatively on young children’s social and emotional development; 11 responses (8.8%) somewhat agreed; 9 responses (7.2%) somewhat disagreed; and 19 responses (15.2%) strongly disagreed and argued that the devices played a positive impact on children’s social and emotional development. There is 64 responses (51.2%) showed no comment/opinions on children’s social and emotional development. The strongly agree and somewhat agree responses were added and categorised into “no-technology” responses; while the strongly disagree and somewhat disagree responses were added and categorised into “go-technology” responses (see Table 1). There was no significant difference of number between the “no-technology” (26.4% out of the total 125 responses) and “go-technology” responses (22.4% of the total 125 responses).

Table 1: No-technology and go-technology responses.

<table>
<thead>
<tr>
<th>Category</th>
<th>Online responses</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-technology</td>
<td>Strongly agree with negative impact</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Agree with negative impact</td>
<td>11</td>
</tr>
<tr>
<td>Go-technology</td>
<td>Disagree with negative impact</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree with negative impact</td>
<td>19</td>
</tr>
</tbody>
</table>

The following is an example from “no-technology” responses. “I have seen it first-hand with our son. His interaction with people has gone downhill... He had been allowed to spend way too many hours a day on video games and also hand held devices and then we started noticing his change in behaviour gradually. He has lost interest in communication, which is very sad to see. He is depressed when he is not allowed to have any hand held devices or video games, he is literally down and just empty and void of happiness.”

The child above is experiencing a behavioural problem – no interest in communication; moreover, this child is also having disruptive and anxious behaviour problems while he is not allowed to play hand-held devices. Similarly, another typical example was that “Lindsay” stated her 4-year-old son had behavioural issues and been nonverbal; she was asking whether she should get him an iPad to help him with his speech as she was told to. The online “no-technology” responses supported the opinion that she should not use iPad in assisting her son’s social development, because “communication is an interactive social skill, which is being significantly eroded by technology overuse”.

The figures in brackets represent the number and percentage of the 33 “no-technology” online responses reporting the nature of the impact of technology on children’s social and emotional development, it should be noted that some online responses reported commitment(s) in more than one category:

- “Anxious” (8, 24.2%)
- “Depressed” (17, 51.5%)
- “Attention deficit” (2, 6.1%)
- “Poor self-regulation” or “addiction” (21, 63.6%)
- “Isolated” (15, 45.5%)

622
On the contrary, the online “go-technology” responses are both similar to and different from the “no-technology” responses. They indicated that “children’s mental health diagnosis is not easy and is a tricky business with many flaws... and a process of assessing a clinical presentation, not a matter of determine course”. The ‘go-technology” responses emphasised the role of the parents or educators in using digital technology as important to children’s development, instead of the technology itself.

The following is an example from “go-technology” responses. “there is huge cultural capital and future prospects in being computer literate. There are many professions that require it as a basic skill, not to mention the many positions in that sector itself that are a natural fit for young adults to bring their enthusiasm to. A passion for video games easily translates into learning to write game code, or character animation. The video game industry makes more moneyed employs more people than the movie industry, Hollywood and Bollywood combined. Contrary to what you seem to believe video games have evolved to be complex social forums, and can be a safe forum for a shy child to develop social confidence and personal skills. They also provide a dangerous world-safe place to explore adventurous play. I agree with limits but they should be specifically geared towards meeting the child’s needs individually.”

The figures in brackets represent the number and percentage of the 28 “go-technology” online responses reporting the nature of the impact of technology on children’s social and emotional development, it should be noted that some online responses reported commitment(s) in more than one category:

- “Engaging” (18, 64.3%)
- “Confidence” (15, 53.6%)
- “Interactive tool” (9, 32.1%)
- “Communication tool” (15, 53.6%)

The argument concern between the “no-technology” and “go-technology” responses lies in whether the use of these technologies was the direct cause of children’s developmental delays or negative behavioural issues in children’s development. The “no-technology” responses were supporting the causation, while the “no-technology” responses were against the causation. For example, one “no-technology” response stated “the problem lies in children who cannot make wise decisions due to the fact that their frontal lobes are not fully developed until age 25. Kids generally pick instant gratification over delayed gratification, e.g. eat entire bag of candy now rather than one piece over a week.” The “go-technology” responses argued “half the problems that is listed from technology could easily be from genetic, or parenting problems: like a divorce, spousal fights/arguments, parents could also be giving their children junk food causing obesity and not enrolling them into sports. The list goes on, and to say that technology is causing all of these problems is absolutely out of context”. Three teenager students (2.4%) presented their experiences of using the technologies from a corollary perspective to the potential negative impact of digital technology. “None of us suffer from any of the problems which have been listed, but it is a good group of around 5 of us. Does that make us all ‘Outliers’ in the Correlations which have been cited, or are the Correlations and their evidence simply not examples of repeatable ‘experiments’ (for lack of a better term)?”

Nevertheless, the argument concerns from both “no-technology” responses and “go-technology” responses also revealed that no child is alike and each child will follow a divergent path in order to achieve developmental outcomes.

4. Discussion and Conclusion

This study used 125 responses from Internet to explore people’s opinions towards hand-held devices, and makes four useful contributions to knowledge on the level and nature about the people’s opinions towards the impact of hand held devices on young children’s social and emotional development.

First, it was found the “no-technology” responses highlighted that hand-held devices are feared to be addictive and socially isolating and minimising time spent with families and peers (Bower, 2013; Brown et al., 2012; Finegan & Austin, 2002; McCarrick & Li, 2007; Plowman & McPake, 2013;
Plowman & Stephen, 2003; Struppert et al, 2010; Yelland, 2011). However, the “go-technology” responses were consistent some of the reported benefits including increased persistence, self-confidence and self-esteem within learning (Dowling, 2014; Yelland et al., 2014; Lee, 2015; Plowman & McPake, 2013).

Second, it was found that one of the primary distinctions between the two responses was the level of parental engagement with the child’s use of the hand-held devices. In the “no-technology” response the parent appeared disconnected from the child’s use of technology, in that “he had been allowed to spend", rather in the “go-technology” response, where by the parent is “right there with her, constantly engaged”. Which confirms Plowman and McPake’s (2013) study that if parents use digital technology as a medium to begin conversation and discuss questions or interests the children have, it allows for children’s growth and development.

Third, while it was found that both groups of responses had strong arguments with examples, it emphasised that every child should be treated as a different case and follow a divergent path in order to achieve developmental outcomes. It was consistent with Dodge et al (2002)’s statement that differences in gender, temperament, interests, learning styles, life experiences, culture, and special needs provide the individual differences that make each child unique and that make development difficult to predict.

Last, both groups of responses recognised the need for young children to learn and develop their skills and knowledge with technologies. Both groups also agreed that overuse of digital technologies could cause developmental issues in young children (Woods & Miano, 2012).

This paper has its limitations. All the online responses were from those people who spoke for themselves and who cared to do so online. Moreover, the responses did not cover all the developmental areas of young children, although some developmental areas such as language and social and emotional developmental areas were mainly focused. Based upon this limitation, a further study will be developed by using properly developed questionnaire to investigate the stakeholders’ opinions, including parents and educators, towards the use of hand-held devices in young children’s development. Moreover, parents’ engagement with use of digital technology will be studied in-depth so that strategies of digital technologies could be provided to assist better use of the devices and software and therefore produce more positive impact on children’s social and emotional development.

References


National Association for the Education of Young Children (NAEYC) and Fred Rogers Centre (2012). Technology and interactive media as tools in early childhood programs serving children from birth through age 8. A joint position statement of the National Association for the Education of Young Children and the Fred Rogers Center for early Learning and Children's Media at Saint Vincent College. Available at: www.naeyc.org/files/naeyc/file/positions/PS_technology_WEB2.pdf (accessed 3 September 2014).


Schwartz SL (2011) The role of peer deviance and social support in the development of symptoms of internalizing disorders among youth exposed to Hurricane Georges. Graduate Degree’s Thesis, University of Kansas, US.


Applying Pedagogy to the Design of Software for Helping Students Learn Equation Solving

Daphne ROBSON
*Ara Institute of Canterbury, New Zealand
*daphne.robson@ara.ac.nz

Abstract: In this paper, we describe the development of software designed to help students learn strategies for equation solving. Instructional design principles were followed and two cycles are described. In the first cycle of analysis, design, development, implementation, and evaluation, prototype software was completed. In the second cycle, after analysis of the prototype, Equations2go was designed, developed and implemented ready for evaluation. Pedagogical principles are fully embedded into the design of Equations2go and are the focus of this paper. Equations2go is now available free of charge.

Keywords: Educational software, pedagogy, instructional design, equation solving, flexible thinking.

1. Introduction

For many years, the potential for computers to be used to help students learn has been seen but it is now well recognised that any use of computers or software for learning must be based on sound pedagogical principles. This paper focuses on the design of software for students undertaking study in applied fields such as Computing, Engineering, and Science who need to learn how to use and manipulate mathematical formulae and equations relevant to their field. The software development followed the ADDIE instructional design model with its five components: Analysis, Design, Development, Implementation, and Evaluation.

In the first phase of the ADDIE cycle, the learning needs of adult students with a variety of backgrounds and prior mathematical knowledge were analysed and relevant pedagogies identified. Software was designed and developed using a rapid prototyping principle in which a simplified version of the main software design was developed. This prototype software was then implemented and evaluated to investigate its impact on equation solving strategies of students (Robson, 2006).

In the second phase, results of the first phase were analysed and the software redesigned, redeveloped and implemented ready for evaluation. The focus of this paper is to describe the relationship between pedagogical principles and the design of the software, Equations2go (Robson & Wratt, 2017).

2. Prototype Design

When the prototype software was designed, relevant pedagogical principles were considered and then applied to the design of features of the software. In this section, each pedagogical principle is described followed by the software feature that it led to.

2.1 Emphasis on Strategies

Researchers have observed that many mathematics teachers focus on teaching procedures, but the type of thinking needed to plan a strategy or set of actions to solve an equation is an important part of learning to think mathematically and to solve problems. Star and Madnani (2004) described this type of higher level thinking as ‘flexible thinking’.
When students use the software, they are encouraged to learn to think more flexibly without being distracted by algebraic details as students make decisions about what action to take and the software carries out the procedure needed for that step.

2.2 Multiple Strategies

Learning activities which help students think with the flexibility needed for equations solving include finding and comparing multiple strategies (Star & Seifert, 2006) and thinking about which strategy is best (Star & Madnani, 2004). The ability to use and select from multiple strategies is widely associated with mathematical expertise (Pegg & Tall, 2002; Jonassen, 2000).

To further encourage students to think with flexibility, the prototype software accepts several different strategies for each equation and each strategy is displayed visually with its own stepping stones pattern.

2.3 Informative Feedback

Much research has found that feedback makes a valuable contribution to learning as it provides communication to students after they have performed a learning task (Hattie, 2012). Most feedback is intended to either acknowledge a correct response or provide information about an incorrect response allowing students to learn from both their successes and their errors. It is generally agreed that feedback improves learning if it encourages students to think actively (e.g., Lyster & Ranta, 1997). It is relevant to consider a study of feedback for algebra. Nguyen-Xuan, Nicaud and Gelis (1997) tested different types of feedback and found that for algebra, it should be short, include consequences of errors, give enough information for students to see why their response was incorrect, but allow them to work out the next step themselves.

The prototype software provides feedback at each step that refers to the action that the student chose rather than just explaining what to do. In other words, the feedback relates to their choice of strategy. When a student chooses a correct action, the software provides positive feedback, explains what progress has been made, and displays a working step for the action along with the results of the action. When a student chooses an action that does not progress towards the solution, information about the goal for that step is provided and students can then try again.

2.4 Visual Interface

Research on whether graphics contribute to learning is varied but any graphics must be integral to the topic and the learning, and should be included for a specific reason (Rieber, 1994). Graphics can allow students to use both visual and verbal channels. By providing visual images, it can help students observe patterns and spot relationships (Pfitzner, Hobbs & Powers, 2001). This in turn can help students develop their understanding of relationships, including those in algebra (Hewitt, 2012).

In the prototype software, graphics were used to provide stepping stone metaphors of “take one step at a time” and “leave no stone unturned” with the latter reflecting the students’ search for different strategies. Other visual metaphors are provided by different paths along the stones representing different strategies, feedback flags or signposts providing guidance, and the light in the tree turning on to reward successful choices. The direction of each successful step uses a metaphor of opposite directions for inverse operations. Students were also able to request that a score be displayed. The interface design of a partially completed equation is shown in Figure 1.
3. Analysis of the Prototype Design

Research involving tertiary students using the prototype software (Robson, Abell & Bousted, 2012) included pre and post-tests, surveys, and analysis of electronic data showing which choices students made. Results relevant to each pedagogical principle follow.

3.1 Emphasis on Strategies

Emphasising strategies rather than the details of procedures was very successful for students when the equation was in their “Zone of Proximal Development” as described by Vygotsky (1978). In this zone, students are unable to solve a problem on their own but are able to solve it with guidance. All students who were unable to solve a particular type of equation in the pre-test were able to solve a similar equation in the software and half of these were also able to successfully apply their learning to the post-test and solve this type of equation on their own.

3.2 Multiple Strategies

The majority of students found searching for multiple solution strategies helpful, but the students who made the most use of this feature were those who were already able to solve equations in the pre-test. Some students found multiple strategies confusing, and may have been cognitively overloaded by several strategies.

3.3 Informative Feedback

In the prototype, a quick tip was displayed for each action and students could click to request further feedback. This was well used with logged data showing that students requested the extra feedback an average of three times per minute. Furthermore, students’ survey comments described how useful they found the explanations. The reason for this two stage display of feedback was to keep the main screen as simple as possible (Neilsen, 1993) while allowing students to actively request more information when they needed it (Mason & Bruning, 2001). The majority of students liked the feedback provided by the score as it allowed them to see their progress, but others found it discouraging.

3.4 Visual Interface

The data collected showed no evidence that the visual metaphors contributed to learning. This result supports Reiber’s (1994) assertion that any graphics must be integral to the topic, and suggests that the visual metaphors for abstract principles were not a strong enough reason to use graphics. Although in the trials, students were asked to explore multiple strategies, the software did not make this clear. A need was identified to redesign the interface so that students would be encouraged to explore multiple strategies and hence have the opportunity to increase their understanding.
4. Redesign of the Software

As the emphasis on strategies in the prototype was so successful in enabling students to solve equations they couldn’t solve on their own, this was still the main principle on which the redesign was based.

To encourage more competent students to learn from exploring multiple strategies, but without confusing beginning students, a system of stars was included in the design that recorded the number of ‘important’ strategies found. Important strategies were considered by the author, in consultation with her colleagues, to add to students’ understanding. Each equation began with one or more outlines of stars and these were filled in when each important strategy was found. Other strategies were also accepted by the software and students received other positive feedback for these.

As students requested the extra feedback so often in the prototype and reported favourably on its usefulness for learning, it was decided to display it after every action. To continue to keep the main screen as simple as possible a clean simple interface was designed, without the graphics that had not appeared to contribute to students’ learning. The redesign of the interface is shown in Figure 2.

![Figure 2. Partially solved equation in Equations2go](image)

5. Development and Implementation of Redesigned software

Equations2go was developed as a web application that can be run on computers and mobile devices using a browser. There are two parts to the software: the web application code written by Matthew Wratt using NPM, Webpack and React, and the code for each equation written by the author of this paper using a custom DSL (domain-specific language). The DSL is parsed and compiled into the web application.

The software was first used with students at Ara Institute of Canterbury during the first semester of 2017 and is now freely available at [http://equations2go.ara.ac.nz](http://equations2go.ara.ac.nz). It was originally designed for adult students but it is expected that school students studying algebra will also find it useful. It is hoped that in return for using Equations2go free of charge that users will use the feedback button to provide useful comments that the author can use to improve the software.

6. Evaluation of Redesigned Software

For the evaluation part of this second phase of the ADDIE cycle, We will initially involve tertiary students at Ara Institute of Canterbury who are studying mathematics as part of their preparation for studying technical qualifications in engineering, science, medical imaging or computing. As in the first phase (Robson, Abell & Boustead, 2012), the participants will answer a pre-test and a post-test before and after using the software. This will allow us to see any impact of the software on students’ ability to solve a variety of types of equations. Participants will also complete a survey which will give us more information about how the software impacted on students’ learning and will provide us with information for improving the software.
7. Conclusion

This paper describes a case of using pedagogy as the basis of a software design which was then developed using the instructional design principles of cycles of analysis, design, development, implementation and evaluation. In the first cycle, each pedagogical principle is described along with the software feature that it led to. The prototype was developed and evaluated by students who were studying equation solving. In the second cycle, the results of the evaluation were analysed and used to inform the design of the main version of Equations2go. With development and implementation of the redesigned software complete, we are ready to evaluate it and thus complete the second cycle. As it is important that educational software is designed using pedagogical principles, the contribution of this paper is the description of the strong relationship between pedagogical principles and software features.

Acknowledgements

Many thanks to the software developer, Matthew Wratt, who so willingly developed the main code for Equations2go and who consistently thought of many creative and innovative solutions to my many requests. I particularly appreciate the way he designed and wrote the custom DSL so that I have full control over creating the equations and determining the actions and feedback at each step.

References


630
The Development of Mobile Learning CPD Modules to Improve the Management of Respiratory Diseases

Georgina ORSBORN\textsuperscript{a*}, Teresa DEMETRIOU\textsuperscript{b}, Kerri ARCUS\textsuperscript{a} & Elizabeth ASBURY\textsuperscript{a}

\textsuperscript{a}School of Health, Whitireia New Zealand, New Zealand
\textsuperscript{b}The Asthma and Respiratory Foundation NZ, New Zealand

*Georgina.Orsborn@Whitireia.ac.nz

Abstract: The social and economic cost of Asthma and Chronic Obstructive Pulmonary Disease (COPD) in New Zealand is increasing. In response, new and updated Adult and Child & Adolescent Asthma Guidelines were created by the Asthma and Respiratory Foundation NZ to inform healthcare providers of current best practice. Primary healthcare providers are at the forefront of asthma management and treatment, but the New Zealand population is geographically disparate, and facilitated training can be impractical. Whitireia New Zealand (Whitireia) and The Asthma and Respiratory Foundation NZ worked in collaboration to develop four online interactive modules designed specifically for the primary healthcare sector.

Method: The Asthma and Respiratory Foundation and Whitireia established the intended learning outcomes and vision for the packages. New and existing clinical content was reviewed and an outline, storyboard and prototype for four modules was created. All four modules were initially created using Articulate Storyline Two, and later upgraded to Articulate 360 and were accessed via the CPD@Whitireia Moodle Learning Management System (LMS). Modules were piloted and evaluated by regional healthcare providers using Kirkpatrick’s Four Levels of Evaluation model, with evaluative data recorded on completion of the modules and at six-week follow-up. Sixteen primary healthcare nurses participated in the pilot and evaluation. There were significant increases in knowledge relating to all learning outcomes for all four modules (p<0.01). Respondents reported positive experiences with the content of all four modules, with many aspects gaining a 100% satisfaction rating. Similarly, participants found all aspects of all four modules broadly engaging, with many interactive features gaining a 100% satisfaction rating. On follow up, all participants reported that the modules had provided a substantial positive long-term impact on their asthma and COPD patient care. Evaluative pilot data from the Asthma and COPD modules indicates a very positive response from all participants, in terms of learning, experience and engagement with all four modules. Knowledge aligned to the learning outcomes increased significantly following module completion, with follow-up data illustrating that interactive, well designed eLearning CPD modules can help inform and improve patient care.

Keywords: Mobile learning, Continuing Professional Development, Articulate 360, Respiratory disease.

1. Introduction

Respiratory disease continues to make a substantial contribution to New Zealand’s health burden. The recently released Impact of Respiratory Disease in New Zealand: 2016 Update (Telfar & Barnard, 2016), commissioned by the Asthma and Respiratory Foundation NZ, reported that respiratory disease accounted for one in 10 overnight hospitalisations. It also identified that over 521,000 people take medication for asthma, and over 35,000 New Zealanders are estimated to be living with severe Chronic Obstructive Pulmonary Disease (COPD – long-term lung disease). The report also highlighted the growing social and economic disparities among asthma and COPD sufferers, and that primary care healthcare providers were at the forefront of asthma management and treatment. This prompted the development of the Adult and Child & Adolescent Asthma Guidelines, created by the Asthma and
Respiratory Foundation NZ to provide health professionals with current best practice guidance for asthma and COPD management and treatment.

The Asthma and Respiratory Foundation NZ have provided foundational education for health professionals for many years. Changing educational technologies along with these updated clinical guidelines, prompted a review of how this education was delivered. Therefore, to disseminate this clinical information to all healthcare providers, The Asthma and Respiratory Foundation NZ, in collaboration with Whitireia, identified that removing barriers from undertaking professional development would be optimally addressed by the creation of online continuing professional development (CPD) packages. Traditionally, asthma and respiratory CPD courses have been two-day face-to-face workshops, necessitating both travel costs and time away from work, as New Zealand is a geographically diverse location, with many rural communities having limited access to local tertiary education and training. Therefore, in order to upskill the professional workforce within the community, it is essential that quality professional development is available and accessible for all.

The ability for online education to be accessible via any device is becoming increasingly important. Research NZ (2015) found that approximately two-thirds of all adult New Zealanders own or have access to three or more internet enabled devices, with most preferring their smartphone. Similarly, an Australian study in 2014 found that 87% of students had a smartphone (Rung, Warnke, & Mattheos, 2014). A 46% increase in use of smartphones for adult New Zealanders in last three years (Research NZ, 2015), emphasizes the need for tertiary educators to be providing content that is responsive to any device, desktop, tablets and smartphone alike. If designed well, tertiary educators can then create activities that encourage active eLearning by providing learners with engaging interactive resources which require learners to become actively involved in their learning and to reflect on their actions (O’Donnell, Lawless, Sharp & Wade, 2015). This project maximized the strengths of both partners: the clinical education expertise of Whitireia and the research and leadership in the clinical management of asthma and COPD of the Asthma and Respiratory Foundation NZ.

Therefore, in order to evaluate the accessibility and educational value of the four Asthma and COPD modules, a pilot evaluation was undertaken among a target population of primary healthcare nurses who work at the frontline of asthma and COPD management, using the Kirkpatrick Learning and Training theory (Kirkpatrick & Kirkpatrick, 2006). This paper presents the evaluation data from the four eLearning modules.

2. Methods

2.1 Modules Development and Content

The development team, consisting of educators from both The Asthma and Respiratory Foundation and Whitireia, initially met to establish the intended learning outcomes and vision for the packages. New and existing clinical content was reviewed and an outline for four modules was created. From there a storyboard and prototype of the first module was developed and then reviewed by the team. All four modules were initially created using Articulate Storyline two, and later upgraded to Articulate 360 (Articulate, 2017) and were to be accessed via the CPD@Whitireia Moodle Learning Management System (LMS). Rapid authoring tools such as Articulate 360 allow non-programmers such as tertiary educators to take existing course content and create dynamic online educational material (O’Donnell et al, 2015). The modules were not outsourced to specialist eLearning services to be developed, but were instead developed by postgraduate educators with both clinical, educational and eLearning development expertise. The ability to simultaneously ‘walk the line’ between eLearning developer and academic, enabled the developer to interpret complex clinical content and create engaging interactions and scenarios that were meaningful to health professionals, without repetitive contact. Every stage of development was reviewed in partnership with the subject matter experts (SMEs) from the Asthma and Respiratory Foundation, often remotely using Articulate Review.

Articulate Storyline 360 is also SCORM (Sharable Content Object Reference Model) compliant, enabling tracking of participant’s progress and completion results on LMS. Storyline 360 allows developers to create content that is made more engaging by including dial and slider interactions, video and audio interactions, quizzes, hotspots, variables and branching scenarios. Most importantly...
Storyline 360, allows developers to create responsive course content that can be viewed anytime, at any place and on any device. Accessibility to education was deemed critical in the creation of these courses. The result was the creation of four interactive and self-paced modules that enabled participants to demonstrate new knowledge and apply learning. The modules were designed to be completed within eight to twelve hours in the following order: Asthma Fundamentals, Asthma Management, COPD Management, and Health Promoting Practice. Each module addresses three learning outcomes and were designed for healthcare professionals in all settings. Modules were reviewed by Health Literacy New Zealand to ensure clarity, along with clinical experts to check that the adaptation of the new guidelines were accurate. On completion, the modules were reviewed and tested internally and then externally to address technical issues, content and relevance to practice.

2.2 Evaluation

The evaluation was based on Kirkpatrick’s Four Levels of Evaluation model (Kirkpatrick & Kirkpatrick, 2006), which proposes a four level training evaluation approach. The first two levels, Reaction and Learning can be assessed on completion of the course, and include both the participants perceptions of the training and their perceived increase in knowledge. Level Three explores the extent to which the new learning has been applied on the job, and requires post-participation follow-up. Level Four evaluates the organizational impact of the training (for example, improvements in patient outcomes), which fell outside the remit of this investigation.

2.2.1 Participants and Procedures

Participants were recruited from a cohort of primary health care nurses working in a geographically rural area of the lower North Island, New Zealand. The nurses were approached to participate in the evaluation by email, which outlined the background of the module development along with the course content. Potential participants were informed that they would be required to undertake all four modules within a six week period and provide feedback using an online questionnaire on completion of each module. Questionnaires were an adapted version of the Whitireia Faculty of Health standard programme evaluation tool, and were piloted internally to ensure clarity and practicality. Changes in understanding based on the learning outcomes were charted on a seven-point numeric Likert scale. Participants were also asked to rate their experiences of undertaking the modules using a four point Likert scale, from ‘Strongly disagree’ to ‘Strongly agree’, and to rate their engagement with the modules using a five point Likert scale, from ‘Not at all engaging’ to ‘Very engaging’. Additional questions were designed to explore mobile device type, internet connectivity, web browser platform and performance issues. A follow-up online evaluation questionnaire was sent to all participants six weeks following completion of the four modules, with seven statements related to learning outcomes and practice rated on a seven-point numeric Likert scale (‘Very much/often’ to ‘Not at all / rarely’). Questionnaires presented on completion of the modules were designed to incorporate Kirkpatrick Levels One and Two, and the follow-up questionnaire reflected Level Three of the Kirkpatrick model.

2.2.2 Statistics

The data were analysed using SPSS. Statistical significance level was set at p<0.05. The study was performed as a within-subjects design, using descriptive statistics and Paired T-Tests with means to find direction. Data are expressed as mean ± standard deviation.

3. Results

3.1.1 Participants

Sixteen primary healthcare professionals agreed to participate in the pilot evaluation. The study population was comprised of 11 (69%) practice/district nurses, three (19%) specialty nurses and two
(12%) nurse educators. Participants had been in their current clinical role for an average of 8.2 (± 6.9) years. Participants reported that each module took an average of 118 (± 54) minutes to complete. Fifty six percent of the cohort completed the modules by laptop, 37% by PC and 7% by tablet.

3.1.2 Module Learning Outcomes

Participants reported a significant improvement in their understanding aligned with the learning outcomes for all four modules on completion of the eLearning packages. Module One learning outcomes; prevalence and causes of asthma in New Zealand (4.75 ± 0.85 vs. 6.37 ± 0.05, p<0.001), anatomy and physiology of respiratory system (5.43 ± 1.03 vs. 6.50 ± 0.51, p<0.001) and diagnostic tools (5.12 ± 0.95 vs. 6.31 ± 0.71, p<0.001) showed significant improvements in knowledge following module completion, while Module Two learning outcomes; approaches to asthma management (4.43 ± 1.26 vs. 5.93 ± 0.92, p<0.001), creating an asthma self-management plan (4.31 ± 1.25 vs. 6.06 ± 0.68, p<0.001) and asthma medication 4.25 ± 1.23 vs. 5.93 ± 0.68, p<0.001) showed similar gains. Module Three learning outcomes; risk factors and incidence of COPD (4.93 ± 1.03 vs. 6.33 ± 0.72, p<0.001), signs and symptoms of COPD (5.20 ± 1.14 vs. 6.53 ± 0.63, p<0.001) and COPD management options (4.86 ± 1.35 vs. 6.40 ± 0.82, p<0.001) had comparable improvements to the Module Four learning outcomes; social determinates of respiratory health (5.21 ± 0.97 vs. 6.64 ± 0.49, p<0.001), health literacy (5.21 ± 1.05 vs. 6.64 ± 0.63, p<0.001) and culturally responsive care (5.71 ± 0.99 vs. 6.57 ± 0.64, p=0.001).

3.1.3 Experience

Participants were asked to rate their experiences of undertaking the modules using a four point response scale, from ‘Strongly disagree’ to ‘Strongly agree’. The table below (Table One) indicates the percentage of respondents who responded positively towards the modules.

<table>
<thead>
<tr>
<th>Module</th>
<th>I found the module was engaging</th>
<th>I thought the content was presented clearly</th>
<th>The module content was relevant to my job</th>
<th>Completing the module has improved my understanding of asthma care</th>
<th>I am confident that I can apply this learning to my job</th>
<th>I was comfortable with how long the module took to complete</th>
<th>I found the module easy to navigate</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>100%</td>
<td>100%</td>
<td>93%</td>
<td>100%</td>
<td>93%</td>
<td>100%</td>
<td>93%</td>
</tr>
<tr>
<td>Two</td>
<td>100%</td>
<td>100%</td>
<td>93%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Three</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>87%</td>
<td>100%</td>
<td>100%</td>
<td>79%</td>
</tr>
<tr>
<td>Four</td>
<td>86%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>79%</td>
</tr>
</tbody>
</table>

3.1.4 Engagement

Participants were asked to rate how engaging they felt each interaction type was using a five point Likert response scale, from ‘Not at all engaging’ to ‘Very engaging’. The table below (Table Two) indicates the percentage of respondents who responded positively towards each module.

3.1.5 Follow-up

At six week follow-up, participants rated seven statements related to learning outcomes and practice on a seven-point numeric Likert scale (‘Very much/often’ to ‘Not at all / rarely’). All participants reported
that the modules had provided a substantial positive long-term impact that informed their asthma and COPD patient care. Responses indicated improvements in all targeted learning outcomes in relation to their practice (see Figure One).

Table Two: Percentage of respondents indicating they found the modules interactions either ‘quite engaging’, ‘relatively engaging’ or ‘very engaging’, with parentheses representing the percentage of respondents indicating the two highest scores.

<table>
<thead>
<tr>
<th></th>
<th>Module One</th>
<th>Module Two</th>
<th>Module Three</th>
<th>Module Four</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drag and drop</td>
<td>100(57)%</td>
<td>100(67)%</td>
<td>93(27)%</td>
<td>100(57)%</td>
</tr>
<tr>
<td>Click and reveal</td>
<td>93(32)%</td>
<td>100(57)%</td>
<td>93(54)%</td>
<td>100(57)%</td>
</tr>
<tr>
<td>Scenario</td>
<td>100(49)%</td>
<td>100(57)%</td>
<td>100(60)%</td>
<td>93(36)%</td>
</tr>
<tr>
<td>Dials and sliders</td>
<td>81(50)%</td>
<td>93(63)%</td>
<td>72(40)%</td>
<td>72(36)%</td>
</tr>
<tr>
<td>Mix and match</td>
<td>93(38)%</td>
<td>100(50)%</td>
<td>100(60)%</td>
<td>100(65)%</td>
</tr>
<tr>
<td>Reflective questions</td>
<td>87(38)%</td>
<td>100(50)%</td>
<td>100(74)%</td>
<td>87(65)%</td>
</tr>
<tr>
<td>Video</td>
<td>70(50)%</td>
<td>81(50)%</td>
<td>72(54)%</td>
<td>100(50)%</td>
</tr>
</tbody>
</table>

Figure One: Mean responses to the impact of undertaking the asthma modules on practice.

4. Discussion

The pilot evaluation data from the four Asthma and COPD modules created by The Asthma and Respiratory Foundation NZ in collaboration with Whitireia New Zealand indicates a very positive response from all participants in terms of both learning, experience and engagement. Analysis of the participants’ knowledge measured against the learning outcomes following completion of the modules showed significant gains in comparison with their own reported pre-module understanding. Follow-up data also indicated that undertaking the modules had an appreciable positive long-term impact on their clinical practice, informing and improving their approaches to patient care.

The modules were completed on a number of electronic devices with few technical complications. However, the results relating to engagement with video may be associated with reported web-based connectivity issues. Participants found the interactive interface of the modules highly engaging, with ‘drag and drop’, ‘click and reveal’ and scenario based questions gaining almost unanimous satisfaction.

This pilot evaluation has demonstrated that complex healthcare teaching can be provided through an online eLearning CPD platform. Hernam and Mustea (2016) believe that interactive eLearning material helps articulate important course content and encourages reflection and integration of the new information. Most importantly interactive material is thought to help learners remember core concepts as content is presented in a number of ways to suit varied learning styles. This places the
learner in the centre of the experience (Bergmann & Sams, 2012), while the online modules can be completed ‘anytime - anywhere – any pace’ on any mobile device. Participants can therefore review or revisit content as often as they want to consolidate knowledge.

The vision promoted in the New Zealand Health Strategy is for all New Zealanders to ‘Keep well - stay well - get well’ (Minister of Health, 2016). Nurses and health professionals are the front line promoting optimal respiratory health for everyone, supporting people who live with the long-term conditions of asthma and COPD to live and stay well, and preventing serious problems that lead to loss of enjoyment of life and hospitalisations. It is crucial for health professionals to have access to current, accessible, research-based professional development that is fit for purpose. This successful partnership between Whitireia and the Asthma and Respiratory Foundation NZ produced an innovative online continuing professional development course, extending the boundaries of traditional teaching for complex clinical skills. Early evaluation indicates that the modules are relevant, useful and already impacting the knowledge and skills of those who support people who live with asthma and COPD.

5. Conclusion

eLearning modules offer tertiary educators a way to provide engaging, interactive and comprehensive CPD education, with the flexibility required to suit busy healthcare professionals. There are a number of challenges, including the invested time required to master the authoring tools, but the rewards are rich and the potential to improve learner engagement without geographic limitations are extraordinary.

Acknowledgements

The authors would like to thank Renee Catherall for her help with Moodle and data collection, and to Wendy Trimmer for her support.

References

Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. International Society for Technology in Education; USA
The Effectiveness of Media Platforms on Reading Comprehension: A Meta-analysis

Bing Xu*, Guang Chen*, Yuting Suna & Ronghuai Huangab
aSchool of Education Technology, Faculty of Education,
bSmart Learning Institute,
Beijing Normal University, China
teastick@gmail.com

Abstract: The debate about whether digital devices are better than paper to support reading has been last for decades since e-book was introduced. Many researches have focused on this issue and drawn different results. The present study aims to synthesize them by meta-analysis method and gain a comprehensive conclusion. This paper analyzed data extracted from 27 published studies, concluding that PC and iPad have a positive influence on reading comprehension but other digital devices are not. In addition, text genre and the grade of participants are important moderators.

Keywords: media platforms; reading comprehension; meta-analysis

1. Introduction

The term “e-book” is used to describe a text or a book, which is digitally displayed on the screen of a computer, a personal digital assistant, or a specifically designed reader. It may comprise text, graphics, video, animation, and sound (Jeong, 2012). Reading from the computer screen and tablet devices (e.g., iPad and Kindle), whether for personal enjoyment, business, or education is becoming progressively common (Hermena et al., 2017). Compared with the paper books, electronic books have many advantages. They can provide us extra reading experience, such as audio/video playback and hyperlink, which means a better interaction with readers. Digital devices are more portable, economic, environment-friendly and benefit for fragmented reading because of their traits of portability and high-capacity. Readers can highlight the sentences with colors, take notes as well as undo such changes on e-books. What’s more, its agreeable price wins much preference.

Plenty of e-books have emerged since the first one was introduced in the late 1970s during Project Gutenberg and published on the internet (Hart, 2017). More and more people tend to purchase e-book so that its sales volume surpassed paperback in January 2011 according to the data from Amazon (Yarow, 2011). However, the rapid growth disappeared and the popularity of paper turned back from 2015 (Preston, 2017).

Jones and Brown (2011) emphasized reading skills are critical for academic and personal growth. Nevertheless, it’s still a prevalent topic to determine which one is better for reading in the forty years when the electronic book coexisted with the physical book.

2. Literature Review

Whether people are able to read and comprehend information effectively from digital media is a hot topic since the arrival of personal computers in the 1980s (Ball & Hourcade, 2011). There were some investigations comparing the effectiveness of media platforms on reading, and the debate of reading effectiveness among mobile devices, paper, and computer screen is still ongoing (Grimshaw, Dungworth, McKnight, & Morris, 2007; Jeong, 2012; Margolin, Driscoll, Toland, & Kegler, 2013; Sackstein, Spark, & Jenkins, 2015).
2.1 Reading Comprehension

Reading is a complex process and its assessment is also difficult. As for reading outcome, Church regarded comprehension as the measurement of reading process. Baker (2010) also explained that comprehension is an essential factor when accessing reading effectiveness. Though it’s complicated to monitor the reading process, researchers arrived on one common proposition “the purpose of reading is comprehension” (Farr & Carey, 1987). Thus, comprehension is clearly a necessary clue that the current research should follow, which was always indicated by the scores of reading test in former investigations.

2.2 Effectiveness of Different Media Platforms on Reading Comprehension Performance

Mixed conclusions are found among previous studies (Ackerman & Lauterman, 2012; Korat, 2010; Masataka, 2014; Sun, Chieh, & Huang, 2013). According to Korat, e-books played a more positive role than paper books. And Alfassi (2000) also illustrated e-books could assist the learner to deduce, solve problems, and construct learning strategy. However, other researches showed people’s digital reading comprehension score was significantly lower than the paper reading (Ackerman & Lauterman, 2012; Jeong, 2012). Thus, we need to use vast data and a more objective method to explore deeply.

2.3 Other Factors

Reader’s age and familiarity toward devices are key elements mentioned before (Ball & Hourcade, 2011; Cheng, Zheng, Li, & Chen, 2014). Ball and Hourcade (2011) indicated that age was statistically significant on comprehension when comparing reading from paper and computer. Chen et al. (2014) pointed digital reading performance was relevant with reader’s familiarity towards the devices. Readers’ age sometimes has a correlation with familiarity. Additionally, different literature types would trigger different information process activities, inference, and integration (Mo, Wang, & Leng, 2012), so genres of literature should taken into consideration (Yoo, 2015). Overall, the impact of such elements must be considered when investigating present topic.

2.4 Research Questions

Based on the theory of reading comprehension and previous studies about this issue, the current research focuses on two questions:

Q1: Are the digital devices, including mobile devices (e.g., iPad and Kindle) and computer screen useful to support reading comprehension performance?

Q2: How age and text genre moderate the impact of digital devices on reading comprehension?

3. Method

The present meta-analysis follows the steps of formulating the research problem, searching and collecting of studies, coding, data analysis, constructing results and interpretation (H. Cooper & Hedges, 1994).

3.1 Study Source

For relevant published studies from January 2007 to March 2017, a search on Web of Science, MEDLINE, SciELO Citation Index, and ERIC was conducted. The search included terms (key words) for the digital readers (including e-book, e-reader, computer screen, computer-based reading, tablet, electronic devices and media platform) and reading comprehension (including reading, reading comprehension, comprehension, the effectiveness of comprehension).

The inclusion and exclusion criteria used in this study are as follows:

- Studies were published in English and from January 1st 2007 to March 1st 2017.
As many other meta-analysis papers, the studies must be journal papers (Brydges et al., 2015; Hatala et al., 2014; Shin, Park, & Kim, 2015). Conference paper and academic dissertation were excluded.

Studies must include the experimental and control group, compare the effectiveness of paper and one of these two kinds of digital methods, and demonstrate the score of a reading comprehension test.

Finally, after selected by two researchers, a literature pool of 16 journal papers including 27 studies was constructed.

3.2 Data Extraction

According to the literatures of Lipsey and Wilson (2001), Copper (2009), and Pearson (2005), a coding framework containing 33 variables was developed. Two coders coded all 27 studies, and interrater reliability was generally high, with \( \kappa = .747 \) (Landis & Koch, 1977). Uncertainty and disagreement were resolved through discussions among coders. Eventually, the data were inputted into the software Stata 14\(^\circ\) to calculate effect sizes.

3.3 Data Synthesis

For each study, we calculated the standardized mean difference (SMD, Cohen’s d effect size) with mean and standard deviation (SD) (Cook et al., 2011). The presented study chose random effects model because the effect sizes were diversified (Borenstein et al., 2009).

4. Results

The search result included 151 relevant papers. After screened (Figure1) by researchers, 16 journal papers including 27 quasi-experimental or experimental studies matched criteria. These researches covered 2113 participants aged 4-69 years old, from K-12 school, university and community schools.

4.1 Overall Effect Size

As the forest plot (Figure2) shows, the overall effect size illustrated electronic media platforms had a slight positive effect than paper (SMD=0.032; \( p=0.605 \)) (Cohen, 1988). And these studies showed moderately consistent results (I\(^2\)=46.9%) (Higgins, Thompson, Deeks, & Altman, 2003).

We found 19 comparisons between computer and paper (SMD=0.07; \( p=0.411 \)). The pooled effect size was small and nonsignificant. The inconsistency was high (I\(^2\)=57.6%). There were 6 studies about iPad VS paper, of which the pooled effect size was also slight (SMD=0.071; \( p=0.401 \)). These
results indicated that iPad and computer have a positive but nonsignificant influence on reading than paper books. On the contrary, the comparison of Kindle VS paper revealed the negative and nonsignificant influence of Kindle (SMD=-0.228; p=0.213).

4.2 Subgroup Meta-analysis

4.2.1 Text Type

Yoo (2015) and Mo (2012) have illustrated the genre and difficulty of content could influence reading performance. This paper divided 27 studies into different groups according to text genres. As it’s shown in Table 1, people would comprehend better when using digital devices if they read narrative (SMD=0.136, p=0.171) and practical essay (SMD=0.117, p=0.142), however, the results turned to the opposite when it came to expository essays (SMD=-0.163, p=0.055). All the p-values were more than 0.05 which meant the differences were not significant.

Table 1: Subgroup meta-analysis according to text type

<table>
<thead>
<tr>
<th>text type</th>
<th>effect size</th>
<th>weight (%)</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>narrative essay</td>
<td>0.136</td>
<td>55.75</td>
<td>(-0.059, 0.331)</td>
<td>0.171</td>
</tr>
<tr>
<td>expository essay</td>
<td>-0.163</td>
<td>30.62</td>
<td>(-0.330, 0.004)</td>
<td>0.055</td>
</tr>
<tr>
<td>practical essay</td>
<td>0.117</td>
<td>13.63</td>
<td>(-0.039, 0.274)</td>
<td>0.142</td>
</tr>
</tbody>
</table>

4.2.2 Grade

The age and grade of participants were also discussed before (Ball & Hourcade, 2011; Cheng et al., 2014; Zsofia K. et al., 2014). A pooled effect size was calculated to explore this factor. There was a positive impact of electronic devices than paper if the participants were younger than senior high school students or elder than undergraduates. All the p-values were more than 0.05.

Table 2: Subgroup meta-analysis according to participants’ school phase

<table>
<thead>
<tr>
<th>Grade</th>
<th>K-3</th>
<th>grade 4-6</th>
<th>junior high school</th>
<th>senior high school</th>
<th>undergraduate</th>
<th>more senior</th>
</tr>
</thead>
<tbody>
<tr>
<td>effect size</td>
<td>0.194</td>
<td>0.185</td>
<td>0.026</td>
<td>-0.123</td>
<td>-0.094</td>
<td>0.084</td>
</tr>
</tbody>
</table>

5. Discussion and Conclusion

The current research revealed the effectiveness of different media platforms on the reading comprehension. The overall effect size illustrated only a small and non-significant strengthening of electronic devices than paper (SMD=0.03, p=0.606), which was consistent with former researches (Jones & Brown, 2011; Mangen, Walgermo, & Brunnic, 2013; Sun et al., 2013; Yu, 2010). The absence of significance meant no essential gap between digital method and traditional way. Computer-based reading showed the more advantages of all platforms. This result could be explained by the relationship between reading behaviors and devices’ function. Margolin (2013) pointed following along with a finger or mouse and moving lips were obviously positive for comprehension. What’s more, surfing and hyperlink functions in e-books are also benefit for searching information and forming a clue or map of knowledge. These beneficial factors all existed in computer- and iPad-based books, which were more effective for reading than paper and Kindle.

As for text genre, according to the data, people performed better but not significant when reading narrative essays and practical writings like advertisements, letters and etc. Compared to expository essays, those two genres are easier to comprehend and memory. The working memory taken
up by cooperating digital hardware tended to be more important when people read obscure expository texts (Ackerman & Lauterman, 2012; Jeong, 2012).

As for the grade, there was a decrease of effect size along with the increase of age. This tendency showed reading comprehension ability may be related to the reading habit rather than time. That is to say, though the K-3 children may spend less time on reading by digital devices than the elders, they are digital natives who almost use devices to form reading style. Grimshaw (2007) also found children’s comprehension was positively affected by the computer screen. People older than university students showed better reading performance than expected. It may due to research bias caused by the lack of studies focusing on participants in this category.

Overall, people will gain better reading comprehension when using digital devices if they read narrative and practical writings, or if they are young digital natives. Reading habit and genre can moderate the effectiveness of media platforms on reading performance.

Additionally, there are maybe other factors influencing the result, such as the difficulty of reading comprehension test. A theory named “the levels of comprehension” involves three reading comprehension levels, namely, literal, inferential and evaluative level (Alonzo, Basaraba, Tindal, & Carriuave, 2009). The former researches barely introduced which level they focused on, however, different questions’ level may cause different result. What’s more, as we mentioned before, study-bias may cause the research question of “how age regulates the effectiveness of digital devices on reading” unaccountable, and hindered the present issue to be illustrated more comprehensively and authoritatively. In this context, we expect our team and other researchers in this area to explain the relevant information more specifically, and expand the age range of participants.

References


Explore the Impact of Collaborative Tendencies in the Flipped Classroom on Taking Basketball Teaching

Tosti Hsu-Cheng Chiang  
Graduate Institute of Mass Communication, National Taiwan Normal University, Taiwan  
tosti@ntnu.edu.tw

Abstract: Along with the development of internet and mobile technology, wireless communication expands the scope of communication improve portability and convenience. Student can get into and try to integrate into the learning environment, for validate the benefit of learning. The research based on three on three cooperative learning basketball games that focus on flipped classroom concept and develop one suitable basketball. Different teaching strategies understand the students’ learning collaborative tendency. This research takes 401 students in the central university of Taiwan to participate in the experiment which using a tablet (on pre-class, during the class, and after-class). The teams conduct observed in the classroom every week, and took questionnaire method as empirical data to analysis student’s behavior.

Keywords: collaborative tendencies, basketball teaching, flipped classroom

1. Introduction

The advancement of internet technology and the application has been developing rapidly, and the technology in the 21st century has fulfill human needs, people demand and reliance on information technology become increasingly close, while in the learning field, also receive the impact emerging of technology, opening up more possibilities than traditional teaching methods. The role of technology in education is gradually important from the policy of various disciplines, it is easy to see the trend of globalization. In the field of physical education, through the integration of information, not just can combined health education, but also can induce student’s interest, and provide student to the self-learning environment and cultivate self-learning ability, such can enhance the learning effectiveness and professionalism of physical and health education (Hall, Brajtman, Weaver, Grassau, & Varpio, 2014).

Motor skill learning is the main course of physical education, in the past physical education use direct guidance method, attached importance to the demonstration of action and instructed students to imitate and practice, rarely through mobile devices induce encourage student to try different change of body movement, of course student also difficult to experiment different function of body movement. If the course add information and integrate to the arrangement and the operation of multimedia teaching, understand the process of movement skill, skill of insight, build a concept, and what get from movement’s experiment, can make the learning results achieve more with less (Raiola, & Tafuri, 2015).

The basketball teaching is divide into basic training movement and more advance skill application. While, three on three basketball’s competition regulation is belong to skill application, besides need the right and accurate movement, still need defense and offense of proficiency process, in the face of different condition of teamwork’s defense and offense, it need more advanced and rapid resilience, while it important to spatial balance and their relation in the condition change quickly. In addition, basketball movement is belong to teamwork movement, each teamwork must interworking, and know very well about basketball knowledge and the regulation of competition, thus make people who participate can understand basketball knowledge and regulation, should become familiar important issues (Severini, 2014).

In conclusion, on design of basketball flipped classroom must include deep content and assisted materials by the use of multimedia to understand motor skills, build skill and concept. This design hope through application program, with the different teaching strategies whereby understand the impact of
different teaching strategies to the tendency of cooperation and learning motivation. At the same time, hope student can in-depth understand basketball movement, then enhance learning’s motivation and achieve in-depth cooperation in the learning experiment, and get better basketball experience (Huang, Tu, Wang, Chen, Yu, & Chou, 2017).

This experiment question research as follow:
- Observe the collaborative behavior in the flipped classroom on taking basketball teaching
- The implementation of basketball flipped classroom learning activities, to explore the differences in student cooperation.

2. Literature Review

In the application of human society, internet gradually expand, deepen, there are many different from the past face to face teaching model of learning mode, such as MOOC (Massive Open Online Course, large-scale open online classroom). A new Large-scale free online open course with 5 to 10 minutes of small unit segmentation video course, with occasional online discussion and feedback during the passage, online peer learning and discussion, virtual online experiment and online practice and assessment of the course, students can schedule their studies at their own pace. Flipping classroom is a new line of science and technology, and the current higher education is becoming more personal and adaptable, and through the assessment of practice models to ensure that cohesive learning experience is equally valued. (Wanner, & Palmer, 2015)

2.1 Flipped Classroom

The focus of the classroom is not on whether the teacher has a recording video to teach, but to flip the classroom to create a framework for students to get personalized and suitable learning. Even teachers and students can have more interaction and discussion in the classroom (Sams and Bergmann, 2013). Flipped classroom is not intended to replace the teacher, but to allow students in the classroom. It can be more independent inquiry and collaboration, build peer interaction, the teachers in the classroom can play a guide and the role of assisting, rather than being completely dominant, the flipped classroom is to redefine the role of teachers and students, so it refer to move the problem on traditional teaching method to outside classroom (Lage, Platt, & Treglia, 2000).

The flipped classroom contains many benefits, including the interaction between the teacher and the student, and the practical application of the classroom in many disciplines, such as economics (Roach, 2014) and nutrition studies (Gilboy, Heinerichs, Pazzaglia, 2015), and even paper-making cranes, can indeed lead to better learning outcomes.

Flipped classroom is not intended to replace the teacher, but to allow students in the classroom can be more independent inquiry and collaboration. By peer interaction, the teachers in the classroom can play a guide and the role of assisting, rather than being completely dominant, the flipped classroom is to redefine the role of teachers and students, so it refer to move the problem on traditional teaching method to outside classroom (Lage, Platt, & Treglia, 2000). Thus this research develop digital game based on programming learning system with high interaction than video based teaching strategies, and is use to flipped classroom, and attempting to compare the result differences of using flipped classroom teaching strategies.

2.2 Mobile and Ubiquitous learning

Mobile and ubiquitous learning computing concept is refer to via wireless network, ubiquitous space that allow everyone a small node in a huge computing network that interacts with people on a wireless mobile devices. Thus, mobile and ubiquitous computing are not concern only with individuals or objects, but our interaction with mobile devices, it also can correspond to the life needs, and all information can become a module, through search, mixed and remake to become person information, make the human nature of the technology develop and achieve U-oriented society. Mobile and ubiquitous learning include freedom and convenience, personalized learning, spontaneous learning and sharing the resources and interaction (Hwang, & Tsai, 2011).
2.3 Collaborative Learning

Collaborative learning refers to in order to student understand deeper the knowledge of social construction in the social negotiation process (Dillenbourg, 1999), and through information exchange, get the feedback interaction with peer, together construct knowledge. Collaborative learning train the communication skill and social awareness, and produce different awareness and the ability of solving problem. (Scardamalia, 2002; Stahl, Koshmann, & Suthers, 2006) recent years more research point out that collaborative learning can help to enhance learning achievement (Yang, & Chen, 2012)

3. Research Method and Design

The purpose of this research is to understand the difference between collaborative tendencies, learning motivation and learning achievement in different teaching strategies and contexts of traditional teaching and mobile learning. The course includes four parts: (Basketball Vision), (Basketball Justice), (Basketball Venus) and (Taiwan Fans Meetings). It aims to establish the cognitive ability of learners in common sense, techniques and rules. For three on three basketball the necessary knowledge of a better understanding, it lead to their participation in the motive of basketball. Through the information media and curriculum interaction, student can have to appreciate, review and comment on a variety of the movement of the kind of appearance, in order to exercise, communication, science and technology convergence of the curriculum philosophy, to provide student to diverse learning environment.

3.1 Research Framework

The design of Independent variables is suitable for different teaching strategies. This research will divided into three different classes: (APP flipped classroom group), (APP big screen teaching group), and (traditional teaching group). APP flipped classroom group is allow the student download the APP and shall be prescribed reading first, retain 10 minutes to review at the classroom, immediately get into exercise. APP big screen teaching is make the APP projection into big screen, so that students can learn the relevant knowledge in the classroom before get into exercise. Traditional teaching group be directly oral teaching on the court and help the strategy board, then get into exercise.

3.2 Research Tools and Analysis Method

The research use collaborative tendencies scale method design, proposed by Hwang, Shi, and Chu (2011). which focus on collaborative tendencies to conduct survey, and using Cronbach’s α coefficient as verification scale, the reliability of 0.91. The topic of this research using Likert five scale to give score, 5 points on behalf of very agree, 1 points for very disagree). This research will count on 10 questions when proceed statistics analysis, with maximum score is 50 points.

4. Experimental Design

The participants of the study were students of the “physical education- basketball” course at University in the central region. There were eight classroom participate the experiment among with two classrooms using (traditional classroom), three classrooms using (APP big screen teaching), three classrooms using (APP flipped classrooms). Eight classes are taught by one teacher, the courses aims to take three on three basketball competition as end of term the class, the content of course is all about basketball skills, rules, offensive and defensive courses, all of the class level and teaching time are same. Focused on the participants using teaching strategies then eliminate missing value, the result as shown Table 1.
Table 1: The frequency table of teaching method

<table>
<thead>
<tr>
<th>Teaching strategies</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP flipped classroom</td>
<td>149</td>
<td>37.20%</td>
</tr>
<tr>
<td>APP big screen teaching</td>
<td>149</td>
<td>37.20%</td>
</tr>
<tr>
<td>Traditional classroom</td>
<td>103</td>
<td>25.70%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>401</td>
<td>1.001</td>
</tr>
</tbody>
</table>

5. Experimental Results and Analysis

The purpose of this research is to explore the impact of different teaching strategies on student’s collaborative learning tendencies and teaching achievement. In order to verify the effectiveness of the learning model, this research uses quasi-experimental design. It will bring on experiment participants divided into APP flipped classroom group, APP big screen group (experiment group) and traditional teaching (control group), conduct different teaching learning activities, and through before and after scale test of the collaborative tendencies and the achievement test of three on three competition, analysis the change of before and after test of teaching activities. If experiment group and control group has significant differences, it can infer that APP teaching method, for student’s learning achievement and collaborative tendencies has positive impact, and also as a basis for the development of adaptive teaching system. The confidence level of the research is 0.05.

In order to understand the collaborative tendencies of three groups in the learning activities whether has significant differences or not, will take independent sample T-test of collaborative tendencies to assess the cognitive basis of the three groups before the experiment. The result of the experiment shown as Table 2 shows, the statistics result of the collaborative before the test found that, the average score of APP flipped classroom group is 26.81 points, the average score of APP big screen group is 28.52 points, traditional teaching group is 29.06 points. There is no significant differences between APP flipped classroom group and APP big screen group before the test (t=-0.984, p = 0.326>0.05). There is no significant differences between APP flipped classroom group and traditional teaching group before the test (t=1.172, p = 0.243>0.05), there is no significant differences between APP big screen group and traditional teaching group before the test (t=0.283, p 0.777 >0.05). Based on statistics above, it have not reached significant level, showing that each group of student in the experiment before the experimental collaborative tend to have same basis.

Table 2: The summary pre-test table of Independent sample T-test of collaborative each group

<table>
<thead>
<tr>
<th>Collaborative tend</th>
<th>Experiment</th>
<th>Frequency</th>
<th>Mean</th>
<th>SD</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>APP flipped classroom</td>
<td>149</td>
<td>26.81</td>
<td>15.083</td>
<td>-0.984</td>
</tr>
<tr>
<td></td>
<td>APP big screen teaching</td>
<td>149</td>
<td>28.52</td>
<td>14.947</td>
<td></td>
</tr>
<tr>
<td></td>
<td>APP flipped classroom</td>
<td>149</td>
<td>26.81</td>
<td>15.083</td>
<td>-1.172</td>
</tr>
<tr>
<td></td>
<td>Traditional classroom</td>
<td>103</td>
<td>29.06</td>
<td>14.896</td>
<td></td>
</tr>
<tr>
<td></td>
<td>APP big screen teaching</td>
<td>149</td>
<td>28.52</td>
<td>14.947</td>
<td>-0.283</td>
</tr>
<tr>
<td></td>
<td>Traditional classroom</td>
<td>103</td>
<td>29.06</td>
<td>14.896</td>
<td></td>
</tr>
</tbody>
</table>

In order to compare the collaborative tendencies of each group whether has significant differences or not, focused on post-test result to conduct ANCOVA analysis. To meet the basic hypothesis of covariance analysis, homogeneity test of variance and homogeneity test of regression coefficients in three groups were performed. There is no significant differences between experimental group and control group (F= 1.012, p=0.365 >0.05), shown as Table 3, it means that the slope of the regression lines was same. By two kind test above, it can be learned that the experimental group and control group of the three groups have the homogeneity, consistent with basic assumptions of covariance analysis, can proceed ANCOVA analysis.
Table 3: The pre-test table of two way ANOVA of different teaching strategies

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching strategies *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborative tendencies</td>
<td>373.180</td>
<td>2</td>
<td>186.590</td>
<td>1.012</td>
<td>0.365</td>
</tr>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>72853.195</td>
<td>395</td>
<td>184.438</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After the end of the experiment, three student groups take collaborative tendencies post-test, then analysis each group in the different teaching model, different learning strategies which cause by learning differences. According to the result of one-way analysis of covariance shows, after student use different teaching strategies, the post-test score of collaborative tendencies which after eliminate the pre-test effect, there was significant differences between each group (F= 21.169, p < .05), it shows it produce significant differences for the student collaborative learning after using different teaching strategies. Hence, by the analysis result of the research experiment shown on Table 4, can find that APP flipped classroom student group has significantly higher differences than APP big screen group, while traditional student group also has significantly higher differences than APP big screen group, and there is no significant differences between APP flipped classroom group and traditional teaching group.

Table 4: The summary post-test table of ANCOVA between different teaching strategies and collaborative learning

<table>
<thead>
<tr>
<th>Teaching strategies</th>
<th>Frequency</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP flipped classroom</td>
<td>149</td>
<td>32.27</td>
<td>11.931</td>
<td>32.635</td>
<td>1.116</td>
<td>0.000*</td>
</tr>
<tr>
<td>APP big screen teaching</td>
<td>149</td>
<td>26.62</td>
<td>16.881</td>
<td>26.527</td>
<td>1.113</td>
<td></td>
</tr>
<tr>
<td>APP flipped classroom</td>
<td>149</td>
<td>32.27</td>
<td>11.931</td>
<td>32.635</td>
<td>1.116</td>
<td>0.859</td>
</tr>
<tr>
<td>Traditional classroom</td>
<td>103</td>
<td>32.49</td>
<td>11.844</td>
<td>32.326</td>
<td>1.341</td>
<td></td>
</tr>
<tr>
<td>APP big screen teaching</td>
<td>149</td>
<td>26.62</td>
<td>16.881</td>
<td>26.527</td>
<td>1.113</td>
<td>0.001*</td>
</tr>
<tr>
<td>Traditional classroom</td>
<td>103</td>
<td>32.49</td>
<td>11.844</td>
<td>32.326</td>
<td>1.341</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

6. Conclusion

The research the impact of using different teaching strategy on student’s collaborative learning tendencies and learning achievement. Thus the research try to find out the reverse significant differences and explore collaborative tendencies, learning motivation and learning motivation through APP flipped classroom teaching model, APP big screen teaching and traditional teaching mode.

About the student’s collaborative tendencies, the result of this research shows, APP flipped classroom and APP big screen can help enhance student’s collaborative tendencies, the hypothesis in this research is supported (there is significant differences on different basketball student respectively using different traditional teaching, APP big screen, and APP flipped classroom to the collaborative tendencies). Through different teaching strategies, the collaborative tendencies of student who adopt APP flipped classroom is superior than the APP big screen teaching, while the collaborative tendencies of student who adopt APP big screen is superior than the student who adopt traditional teaching strategies. Advisor point out that because of in the video explaining time will emphasize teamwork, by adopting APP teaching can clearly present collaboration, teamwork action; while using traditional teaching, are more likely to ignore the relevant teaching at the scene.
The teacher reveal, because the flipped classroom can allow students to preview enough to prepare course. Most of the time can be used for discussion and practice, which is a great help for learning achievements; because traditional teaching need tactical board and other teaching tools, plus more practical operation, and whether teachers or students is also more familiar with traditional teaching, and therefore it get good teaching effect. As for the APP big screen teaching, it may be more difficult to take into account the progress of all students, nor if the traditional teaching mode can carry out a lot of practical operation, so whether it is APP flipped classroom or traditional teaching model, the improvement of learning achievement is significantly higher than APP big screen teaching.

This research explores the impact of different teaching strategies on cooperative tendencies, learning motivation and learning achievement. In addition to different teaching model, should consider the personality traits of students themselves, such as pre-prepared knowledge and learning preferences, and provide corresponding ones. Learning strategies, so that in line with the different characteristics of students, to provide students with real assistance, especially in the development of science and technology, new teaching model continuously arise, and more accurate to give students the most suitable learning model.

Acknowledgements

This study was supported in part by the Ministry of Science and Technology in Taiwan under contract numbers MOST 106-2511-S-003 -022 -.

References

Teaching Influence for Perceived Usefulness of Interactive Whiteboard - Based on the Perspective of College Students

Jian Xu*, Zhexu Liu & Peng Zhou
Central China Normal University, China

Abstract: The application of interactive whiteboard (IWB) in classroom teaching is more and more widely. In order to provide reference for teachers' classroom teaching, we measured the perceived usefulness and its influence factor of IWB. Based on the analysis of the IWB classroom instructor characteristics, we set up the students who had had IWB class in Central China Normal University as the research object. The results show that the instruction characteristics, engagement and perceived ease of use have a significant impact on perceived usefulness.

Keyword: Interactive whiteboard; Perceived usefulness; Classroom teaching

1. Introduction

Information teaching equipment is the medium that help students acquire more information in limited area. The IWB is one of the most popular information teaching equipment. IWB is typically considered as a kind of computer connected with the projector. It enhances teaching efficiency in the classroom through sharing information beyond time and space, as well as adaptability. However, most teachers only utilize IWB at the level of displaying the powerpoint and seldom combine the teaching content with multi-form information. Therefore the functions and tools of IWB are not completely used in the teaching process.

In order to make the IWB classroom teaching more optimized, the existing studies mainly suggested the technical characteristics, teaching mode, strategy design, the analysis of teaching process, application effect, application status and so on. These studies are not enough to explore the subjective experience of the IWB teaching from the perspective of the students.

As mentioned previously, we attempted to find the factors which have influenced on perceived usefulness of IWB teaching. Based on the review of the existing literature and the teaching characteristics of the IWB, we chose four factors: students’ perceived ease of use, instructor characteristics, teaching content design and student’s engagement.

2. Theoretical Background

2.1 Perceived Usefulness and Perceived Ease of Use

In 1989, Davis proposed two core beliefs: perceived usefulness and perceived ease of use (Davis, 1989). Davis defines perceived usefulness as the degree to which a specific system improves performance, and defines perceived ease of use as the ease of using the system. He stressed that the external variables (system characteristics, development process, and training) affect the behavior intention through perceived usefulness and perceived ease of use. In subsequent research, different scholars continue to verify and enrich the theory of Davis’s. Pituch and Lee (2006) confirmed that perceived ease of use in the online learning environment can affect perceived usefulness. Lee and Yoon (2009) used the
characteristics of teachers, teaching materials and teaching content design as external variables to influence students' online learning through perceived usefulness and perceived ease of use. This paper focuses on influencing factors for perceived usefulness and perceived ease of use of IWB teaching.

2.2 Perceived Usefulness and Instructor Characteristics

The characteristics of teachers are the degree of care, help and acceptance. Dillon and Gunawardena (1995) confirmed that the three characteristics (teachers' attitudes toward technology, teaching methods and the degree of control over technology) affect the learning outcomes of distance education. According to Webster and Hackley (1997), three teachers' characteristics (information technology ability, teaching style, attitude and thinking) are the factors that influence the success of online learning. Jepsen (2005) demonstrated that the characteristics of teachers have a significant impact on students' performance. Based on existing research, we can summarize that the whole characteristic of teacher applying IWB possibly affect the construction of students' knowledge and experience and the usage situation of IWB. Therefore, this paper assumes that teachers' characteristics can affect perceived usefulness.

2.3 Perceived Usefulness and Student Engagement

Engagement is the degree of interaction between teachers and students in the classroom teaching. From the perspective of constructivism, it is important for students to join social communication to promote the learners' understanding of knowledge and the development of advanced psychological functions. Interaction is a key factor in the success of learners, Martin and Rimm-Kaufman (2015) confirm that the interaction quality of teachers and students can affect the perceived participation. In addition, the quality of interaction can also have an impact on perceived usefulness. For example, Lin (2011) suggested that the frequency of negative critical events in communication would hinder perceived usefulness. Therefore, this paper assumes that interaction affects students' perceived usefulness.

2.4 Perceived Ease of Use and Teaching Content Design

We define learning content design as the degree of learning content teacher have designed meeting the needs of students. Lederer et al. (2000) summarized previous research and put forward that when the website content such as the pictures, tables, text is easy to understand, provide more detailed information, and accurately search the information they need, it can be predicted better perceived ease of use. Lee and Yoon (2009) continued to study and demonstrate that the design of the teaching content has impact on perceived ease of use, and he believes that online learning services will provide accurate and easy to understand learning content, which will facilitate the perceived ease of use. It indicated that the content of IWB classroom teaching according with students' cognitive level and cognitive load will improve the students' perceived ease of use. Through the above theoretical research, this paper constructs the model of this study, shown in Figure 1.

![Figure 1. IWB classroom teaching research model](image)

3. Research Methodology

We carried out an empirical study to verify whether the hypothesis is established, the research process
includes three stages, the preparation of the instrument, data collection and data analysis. A questionnaire was designed to measure the different variables. In order to ensure the content validity, the scale was based on the findings of prior studies and combined with the actual situation and characteristics of IWB classroom teaching. The scale of this study (Table 1) include the 11 observed variables and 5 potential variables, and finally the questionnaire are formed based on scale.

The final questionnaire comprised three parts. The first part is the demographic information, including gender, grade, class and professional. The second part is the scale of perceived usefulness and perceived ease of use, which is based on the study of Davis and contains two items. The third part is the scale of influence factors. It contains the scale of teaching characteristics and teaching content design. These are based on Lee and Yoon’s scale, each of it including two items. The scale of engagement from Martin’s research, which contains a total of three items. The quality attributes was measured using a five-point Likert scale with anchors ranging from “strongly agree” to “strongly disagree.

An empirical study was conducted at Central China Normal University. The teachers had gain training of IWB before and fundamentally use it in the classroom. They display various teaching contents according to their course requirement. So, the questionnaires were distributed to the students who have accepted the IWB classroom teaching. Because the senior students are few have lessons. We choose only freshman, sophomore and junior students investigated. Finally, 200 questionnaires were returned. The valid response rate is 91%.

Table 1: Constructs and items

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Questionnaire items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived Usefulness</strong></td>
<td>(Davis, 1989)</td>
</tr>
<tr>
<td>PU1</td>
<td>IWB enhance my learning effectively</td>
</tr>
<tr>
<td>PU2</td>
<td>IWB improves my learning performance</td>
</tr>
<tr>
<td><strong>Perceived ease of use</strong></td>
<td>(Davis, 1989)</td>
</tr>
<tr>
<td>PE1</td>
<td>On the whole, IWB is easy to use</td>
</tr>
<tr>
<td>PE2</td>
<td>The interaction function of IWB is clear and understandable.</td>
</tr>
<tr>
<td><strong>Instructor Characteristics</strong></td>
<td>(Lee and Yoon, 2009)</td>
</tr>
<tr>
<td>IC1</td>
<td>The instructor provides high-quality instruction through using IWB</td>
</tr>
<tr>
<td>IC2</td>
<td>The instructor delivers instructions clearly through using IWB</td>
</tr>
<tr>
<td><strong>Teaching content design</strong></td>
<td>(Lee and Yoon, 2009)</td>
</tr>
<tr>
<td>LC1</td>
<td>The level of difficulty of the learning contents is appropriate</td>
</tr>
<tr>
<td>LC2</td>
<td>The amount of learning contents is appropriate</td>
</tr>
<tr>
<td><strong>Engagement</strong></td>
<td>(Martin and Rimm-Kaufman, 2015)</td>
</tr>
<tr>
<td>ENG1</td>
<td>Use the IWB during classroom teaching, I am willing to share ideas with other</td>
</tr>
<tr>
<td></td>
<td>students and study materials</td>
</tr>
<tr>
<td>ENG2</td>
<td>Use the double IWB classroom teaching, I am willing to carry on the classroom</td>
</tr>
<tr>
<td></td>
<td>interaction and teacher</td>
</tr>
<tr>
<td>ENG3</td>
<td>In the IWB classroom teaching, I’d like to help other students solve their doubts</td>
</tr>
</tbody>
</table>

4. Results

4.1 Measurement Model

All of the survey participants majored in one of the six disciplines (communication, biology, information management, digital media, physics and history). 30.5% of the survey participants are boys, 69.5% girls, 37% of the survey participants are freshmen, 16% sophomores, 47% juniors. The average of the five construct is about 2.5, indicating that almost half of the students do not acceptance the IWB teaching. The average value of perceived usefulness is the 2.709, which shows that the user's perceived usefulness of IWB teaching is not very good.

SPSS version 22.0 was used to analyze the collected data. Scale validation was done using confirmatory factor analysis. The factor analysis utilized the principal component extraction method and Varimax rotation. It required that factor loadings exceed 0.40. The factor loading values of all
indicator variables are over 0.7, far exceeding 0.40.

Table 2 summarizes factor loadings, Cronbach’s alpha, Composite reliability and Average variance extracted of all indicator variables. Convergent validity exists when factor loadings are greater than the threshold value of 0.50 (Hair, Black, Babin, Anderson and Tatham, 2006) and the average variance extracted (AVE) is at least 0.50 (Fornell and Larcker, 1981). As demonstrated in Table 2, the AVE values ranged from 0.553 to 0.651, thus demonstrating adequate convergent validity. As suggested by Gefen and Straub (2005), AVE is greater than the inter-construct correlations. As shown in Table 3, the AVE values were greater than the square of inter-construct correlations, thus demonstrating adequate discriminant validity. This questionnaire used the Cronbach’s a coefficient to test the internal consistency among items of the same construct. According to Cuieford (1965), a Cronbach’s a value that is greater than 0.7 indicates high reliability. Thus, all constructs can be considered reliable.

Table 2: Reliability and validity

<table>
<thead>
<tr>
<th>Construct</th>
<th>Factor loading</th>
<th>CA</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Usefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU1</td>
<td>.813</td>
<td>.741</td>
<td>.741</td>
<td>.589</td>
</tr>
<tr>
<td>PU2</td>
<td>.799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE1</td>
<td>.799</td>
<td>.740</td>
<td>.769</td>
<td>.624</td>
</tr>
<tr>
<td>PE2</td>
<td>.781</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC1</td>
<td>.863</td>
<td>.772</td>
<td>.788</td>
<td>.651</td>
</tr>
<tr>
<td>IC2</td>
<td>.747</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG1</td>
<td>.809</td>
<td>.784</td>
<td>.787</td>
<td>.553</td>
</tr>
<tr>
<td>ENG2</td>
<td>.714</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG3</td>
<td>.703</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching content design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC1</td>
<td>.751</td>
<td>.747</td>
<td>.713</td>
<td>.554</td>
</tr>
<tr>
<td>LC2</td>
<td>.738</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes.
CA = Cronbach's alpha; CR = Composite reliability; AVE = Average variance extracted

Table 3: Validity

<table>
<thead>
<tr>
<th></th>
<th>PU</th>
<th>PE</th>
<th>IC</th>
<th>ENG</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>.589</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>.257</td>
<td>.624</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>.268</td>
<td>.272</td>
<td>.651</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG</td>
<td>.414</td>
<td>.233</td>
<td>.302</td>
<td>.553</td>
<td></td>
</tr>
<tr>
<td>LC</td>
<td>.348</td>
<td>.376</td>
<td>.328</td>
<td>.313</td>
<td>.554</td>
</tr>
</tbody>
</table>

The number of diagonal bold is AVE value, were greater than the square of Pearson’s correlations

4.2 Regression Analysis of the Research Model

In this study, we tested the research model using monadic regressive analysis and multiple regression analysis. Monadic regressive analysis is the correlation between an independent variable and a dependent variable. Multiple regression analysis is the correlation analysis between two or more independent variables and a dependent variable.

4.2.1 Multiple Regression Analysis for Perceived Usefulness

Table 5 shows the findings of the multiple regression analysis used to examine how well the influence of instructor characteristics, engagement, and perceived ease of use explain students’ perceived usefulness for the IWB. It can be seen from table 4 that the Durbin-Watson value is 2.005, between 1.5 and 2.5, which indicates that the independence of the samples is established, and the value of R square is about 0.478, which indicates that the research model possess moderate explanatory power. The teacher's characteristics was significantly associated with perceived usefulness (p=.014<0.05). The engagement was significantly associated with perceived usefulness (p=.000<0.05). The perceived ease of use was significantly associated with perceived usefulness (p=.001<0.05).
Table 4: An overview of the model

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjust R Square</th>
<th>Std. Error of the Estimate</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.691(a)</td>
<td>.478</td>
<td>.470</td>
<td>.64000</td>
<td>2.005</td>
</tr>
</tbody>
</table>

a Predictors(Constant), Instructor Characteristics, Engagement, Perceived ease of use

Table 5: Coefficient matrix

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>.503</td>
<td>.171</td>
<td></td>
<td>2.938</td>
</tr>
<tr>
<td>Engagement</td>
<td>.495</td>
<td>.070</td>
<td>.453</td>
<td>7.063</td>
</tr>
<tr>
<td>Instructor Characteristics</td>
<td>.169</td>
<td>.069</td>
<td>.163</td>
<td>2.470</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>.194</td>
<td>.060</td>
<td>.204</td>
<td>3.249</td>
</tr>
</tbody>
</table>

4.2.2 Regression Analysis for Perceived Ease of Use

Table 7 shows the findings of the linear regression analysis used to examine how well the influence of design of learning contents explain students’ perceived ease of use for the IWB. It can be seen from Table 6 that the Durbin-Watson value is 1.846, between 1.5 and 2.5, which indicates that the independence of the samples is established, and the value of R square is about 0.376, which indicates that the research model possess moderate explanatory power. The design of learning contents was significantly associated with perceived ease of use (p=.000<0.05).

Table 6: An overview of the model

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjust R Square</th>
<th>Std. Error of the Estimate</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.614(a)</td>
<td>.376</td>
<td>.373</td>
<td>.73162</td>
<td>1.846</td>
</tr>
</tbody>
</table>

a Predictors(Constant), Teaching content design

Table 7: Coefficient matrix

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>1.143</td>
<td>.146</td>
<td></td>
<td>7.820</td>
</tr>
<tr>
<td>Design of learning contents</td>
<td>.623</td>
<td>.057</td>
<td>.614</td>
<td>10.933</td>
</tr>
</tbody>
</table>

5. Discussion

In this empirical study, we analyzed students’ perceived usefulness of IWB teaching. First, we analyzed the relationships between the three constructs (instructor characteristics, engagement, and perceived ease of use) and the one constructs (perceived usefulness). Second, we analyzed the relationships between the design of learning contents and perceived ease of use. From the regression analysis, instructor characteristics, engagement and perceived ease of use are positively related to perceived usefulness. The teaching contents design positively affects perceived ease of use.

The teachers possessing better teaching ability will provide a clear and effective teaching, which will enhance the students' perceived usefulness of IWB teaching. Educational institutions need to provide some course to instructors and need to train them to enhance the comprehensive abilities. Design of learning contents was found to affect perceived ease of use (Pituch and Lee, 2006). In the IWB classroom, teachers transform the abstract knowledge into text, image, sound, video, 3D model
and other forms, consequently increase the students' cognition. When the content displayed through the IWB is very different from student's cognitive experience, students may think that participation is very difficult.

And compared with the traditional teaching, students' participation in classroom can be more improved. According to the result, perceived ease of use has a direct positive impact on perceived usefulness. This results has been verified by many scholars. For example, Lin suggests that perceived ease of use of online learning services has an impact on perceived usefulness (Lin, 2011). Perceived ease of use was found to be a significant antecedent of perceived usefulness (Imamoglu, 2007).

6. Conclusion

Teachers are the main operator of IWB. Most of the previous studies explained teachers’ teaching process and its influence, without examining the students’ perspective. To address this research gap, a research model was developed and tested to explain students' perceived usefulness of IWB teaching.

Learners in the traditional classroom is restricted in terms of time and space, In the informatization IWB classroom, understanding and investigating the perception of students to IWB teaching and its influencing factors are of great importance. By investigating critical factors, our study attempts to fill a gap in the IWB research. Our survey results confirm the research model. Our findings indicate that instructor characteristics, engagement and perceived ease of use are the predictors of the perceived usefulness of IWB teaching. And teaching content design are the predictors of perceived ease of use.

This study contributes to theory and practice in three ways. First, it examined perceived usefulness of IWB teaching from the students' standpoint. Second, it developed a research model to gain a better understanding of students' perceived usefulness of IWB teaching. Third, a longitudinal study was used to test the derived hypotheses.

References


Designing Boundary Activity for Mobile Learning in Science Inquiry

Daner SUN a* & Chee-Kit LOOI b

a Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong
b National Institute of Education, Nanyang Technological University, Singapore
*dsun@eduhk.hk

Abstract: The importance of learning in informal spaces for science education has been increasingly recognized by the educators and researcher across the countries. In this paper, with emphasis on the value of learning in informal spaces, we propose designing “boundary activity” as the knot for tightening the linkage of learning in informal spaces and formal spaces. Based on literature review, the theoretical underpinnings are articulated for conceptualizing the “boundary object” and defining the components of “boundary activities”. The inquiry learning based on the principle of boundary activity is illustrated for informing the curriculum design and implementation in the field of mobile technology supported teaching and learning.

Keywords: Learning in informal spaces, boundary object, boundary activity, mobile technology

1. Introduction

The importance of science learning in informal spaces has been recognized by educators and researchers around the world. Regardless of how they are defined, out-of-school learning experiences have a variety of cognitive, affective, social and behavioral effects that can make a significant contribution to learning (Morag & Tal, 2012). Research findings show that learning experiences in informal spaces can facilitate the acquisition of scientific concepts and the development of inquiry skills, as well as stimulate motivation. Educational documents (e.g. curriculum standards) also endorse teaching and learning practices in informal spaces. However, there is a limited number of successful programs or projects that integrate the merits of learning in informal spaces with formal learning. Nor have the existing programs been rigorously examined. Meanwhile, teacher competency in designing and implementing learning activities in informal spaces further hampers best practice of such learning. The result is an increasing gap between formal learning and learning in informal spaces. Although the ubiquitous use of mobile technology creates various opportunities for connecting the formal learning process with informal spaces, the record of success is limited in terms of curriculum design and implementation. Additionally, there is inadequate documental support (i.e. science curriculum standards, teaching materials, online resources) on best practices for connecting formal and informal contexts with mobile technologies. This constrains the sustainability and scalability of such learning and teaching practices.

To address these issues, the paper will focus on the conceptualization of boundary object and boundary activity for connecting the merits of learning in informal spaces with formal learning, as relates to inquiry-based science learning. A lesson exemplar will be shared with the purpose of illustrating how the principle of boundary activity can be applied in science learning design with the use of mobile technology. The research will be the first one to consider the boundary activity in the field of mobile learning for establishing the connection the learning between formal and informal spaces. In boundary activity, the students’ interactions with informal learning spaces will be the integral part rather than the supplementary part of the standard curriculum. This will transfer teacher and students’ perspectives of learning in the informal spaces, and inform the science instruction supported by technology.
2. Theoretical Underpinnings

The researchers have recognized the importance of getting insights into the mobile learning in the informal context, but challenges on the curriculum design and implementation, as well as the assessment of students’ learning still exist. Jones and two others found that there is little literature that considers the structures needed to support informal and semi-informal inquiry learning (Jones, Scanlon, & Clough, 2013). Mortensen & Smart (2007) points out that although there is a growing effort to create partnerships between schools and informal learning settings, documentation of such projects is limited, and generally reported as examples of “best practice” with little discussion of challenges before or during implementation. In this section, the concept of boundary object is further coined in the field of mobile learning. The principle of boundary activity is defined and discussed. The principle will help the educators and teachers to better fit students’ learning in informal spaces into the formal learning context, with establishing the boundary objects and conduct boundary activities.

2.1 The Conceptualization of Boundary Object

To improve connecting learning in formal space with learning in the informal spaces, boundary object is the proposed as the “knot” for linking the learning in these two contexts. Boundary object, a term which has been discussed in science education. It refers to the common idea generated in scientific work which needs cooperation with among divergent viewpoints and the need for generalizable findings (Star & Griesemer, 1989). It can be either abstract or material, for example, field notes, specimens and artifacts which can be the connections between formal learning and learning in the informal spaces.

In a more elaborated definition from Wenger’s study (Wenger, 1998), boundary object is one type of the connections between communities of practices, namely, artifacts, documents, terms, concepts and other forms of reification and around which communities of practices can organize their interconnections. Tsurusaki, et al., (2012) created “transformative boundary objects” and explored how the transformative boundary objects work in the teaching practices of a teacher with the aim of engaging students in science learning. Three types of boundary objects were discussed: bar graph, research questions and nutrition in the teaching of healthful food. In Gilbert and Priest’s (1997) study, they discussed some external factors for promoting the effectiveness of students’ museum visits and attempted to link their informal learning experience with the topics learned via formal learning. They organized group activities for students to discuss the “critical incidents” in the visits, and meanwhile pre-, during and post activities were designed for students to elaborate their knowledge in and out of the classroom. In this case, the critical incident is the boundary object. In brief, boundary object can be an abstract concept introduced in the classroom and elaborated outside of the classroom, or a guiding question related to a key concept. It can be an event or a science phenomenon which requires students to investigate outside of class and discuss in class. The boundary object can also be a physical object which is generated in or outside of the classroom. In sum, as a metaphor, boundary object is the physical or abstract objects generated by the interaction between boundary of formal and informal spaces, which may play an important role in students’ science learning in various contexts, especially for the outsides activities. Once integrated with the use of mobile technology, there will be more physical representations of boundary objects, for example, concept maps, drawings, photos, videos, notes, etc. and there also will be more opportunities for students to engage in discussing, responding to questions and sharing boundary objects.

2.2 Principles of Boundary Activity

Kisiel (2014) proposed that joining resources from both formal and informal learning settings is an effective strategy that enhances students’ interest in science and STEM (Science, Technology, Engineering and Mathematics) learning. He used a term “boundary activity” to define the activities which connected schools and informal science institutions. The term of boundary activity refers to “those encounters between schools and ISEIs (museums, science centers, aquariums, and the like) that involves some kind of designed program-field trip, outreach, and teacher workshop with specific educational objectives”. In his viewpoint, boundary activity is the mediation for the interaction between the existing/original communities. The connection established by boundary activity is a deeper,
practice-based interaction which has potential to better facilitate interaction between the two communities. The creation of the joint practice-based enterprise is highlighted. It is an enlightened term that enables us to think about what these boundary activities look like in the view of curriculum development? What are the best boundary activities serving for the specific purposes in science teaching and learning? What is the mutual interaction between formal learning and learning in the informal spaces in the boundary activities? Based on the above ideas, we define boundary activity as the learning activities which take place in either formal or informal contexts, and contain at least one boundary object that bridge learning in formal and informal environments.

We continue to get more insights into the findings of the relevant studies for carving the conception of the boundary activity. Rickinson et al. (2004) demonstrated that if field activities were ‘properly conceived, adequately planned, well taught and effectively followed up’, they could offer ‘learners opportunities to develop their knowledge and skills in ways that add value to their everyday experiences in the classroom’. Similar findings were found that when teachers did focus pre-visit preparation of the informal institution, there was an improvement in student learning and attitude (Patrick, Mathews, & Tunnicliffe, 2013; Gilbert & Priest, 1997). Mike Sharples (2014) proposed to employ scripted learning methods to conduct outside inquiry activities. In his study, the teacher initiated a structured activity with the mobile devices inside the classroom, and then each pupil continued the investigation outdoors. Results were then shared, discussed and presented back in class. Falk and Balling (1982) at very early time mentioned that without orientation and preparation, students were more likely to concentrate on non-relevant aspects of the surroundings, rather than those relevant to the learning intended. Specifically, Patrick et al., (2013) thought that field trips need to incorporate problem-solving skills, be tied into the curriculum, focus on the standards, and take into consideration the children’s needs. Thus, with more efforts on designing and implementing boundary activity in structural way, the teaching for learning in informal spaces will be more efficient.

Therefore, we propose three components of a boundary activity: boundary object, structure and learning Objectives based on above literature. We refine the idea to delineate between boundary object, activity structure and learning objectives: (1) The boundary object is a prerequisite for designing the boundary activities. It acts as a knot which serves to bridge learning in and out of the classroom and capture the learning process in the informal spaces. With boundary objects, the boundary activities will better fit into the standard curriculum. (2) Structure: the boundary activity is conducted in the pre-, during- and post-activity pattern to guarantee the continuum and stability of cognition or skills developed across the learning contexts. (3) Learning objective: the learning objectives of boundary activity should be defined based on the curriculum standard and the characteristics of the contextual variables in practice. These three components are proposed to guide the design and implementation of a boundary object. Figure 1 represents the structure of boundary activity and the interaction between formal learning and learning in informal spaces. The boundary activity is conducted across the formal space and informal spaces. It is usually prepared and instructed in the formal learning context (i.e. classroom) prior to carrying out it. With mobile devices, data, evidence and any responses related to the tasks of boundary activities generated in informal spaces can be the representative boundary objects, prompting the boundary interactions taking place.

![Figure 1. The principle of boundary object](image)
3. The Medium for Running Boundary Activities: nQuire-it and WISE

In implementing boundary activity, we envisage the use of a stable and multifunctional system to support students in-classroom activities and out-of-classroom activities, and capture interactions of students and their teacher in the various contexts (i.e. online & authentic learning, in classroom & out of classroom). Two web-based platforms will be integrated in boundary activity based learning: nQuire-it (http://www.nquire-it.org) which facilitates students’ inquiry activities in informal spaces, and WISE (https://wise.berkeley.edu/) which guides students’ online inquiry in a step-by-step manner. Figure 2 represents the overall picture of the roles nQuire-it and WISE play in boundary activity. In comparison of other sensor-based technologies, nQuire-it is a learning platform which is more suitable for conducting outside activities in either guided inquiry or open inquiry, and either in an individual or collaborative way (Llewellyn, 2007; Wenning, 2005). Specifically, it supports students to collect real time data outside (i.e., real experimentation, hands-on activities, home activities, field trips, etc) using Spot-it (a mobile tool for capturing images and annotating things with notes) and Sense-it (a sensor-based mobile tool for collecting and sharing data using phone sensors: accelerometer, gyroscope, light and sound, etc). Thus, the use of nQuire-it will particularly enhance students’ interaction with the informal learning spaces for testing their hypotheses and deepening understanding through authentic learning activities. In this case, the real-time data in the form of photos of scientific phenomena and graphs are the major boundary objects. The activities for planning, conducting data collection and sharing are the boundary activities. As a learning management system (LMS), WISE provides teachers with a powerful authoring tool to design guided inquiry-based activities. With the use of WISE, the learning taking place taking place in formal spaces and informal spaces can be merged into this system. A WISE lesson facilitates students’ online investigation of simulation, videos and virtual labs in and out of the classroom (Slotta & Linn, 2009). The WISE system provides various learning tools, such as a drawing tool, concept map tool, simulations, peer discussion tool and others to support student learning and communication in and out of the classroom. As a result, students receive more opportunities to investigate in the virtual contexts and report their findings online. More importantly, teacher-student and student-student interaction, students’ responses, teacher feedback and assessment can be traced and recorded as evidence of students’ performance. nQuire-it can be flexibly inserted into the learning design of WISE lessons. In this Hong Kong-based study, the synergic use of nQuire-it and WISE enables students to investigate both the virtual and authentic scientific phenomena and to interact with their teachers and classmates any time anywhere.

4. Lesson Exemplar of Boundary Activity based Science Inquiry

Table 1 shows a lesson exemplar based on the principle of boundary activity and inquiry based learning. The topic is Energy, from the Hong Kong General Studies (Primary 6). In the lesson, four inquiry phases consisting of Questioning and Context, Sense-it Exploration, Sharing and discussion, Summary and Reflection are organized by the WISE system and the nQuire-it platform. The step-by-step manner facilitates students’ inquiry learning in an explicit manner. The principle of boundary activity is
integrated in the design of inquiry phases. In the pre-boundary activity stage (Questioning and Context), students are provided with instructions of tasks and guiding questions. During the boundary activities (Sense-it Exploration), students are engaged in a series of hands-on activities with mobile devices: collecting data, uploading data, reviewing data and doing peer assessment. In post-boundary activities (Sharing and discussion), students’ work is further discussed and assessed in class. Finally, they consolidate their understanding and respond to the guided questions in WISE system (Summary and Reflection). Through these kinds of activities, the teacher will have a clearer picture of how a boundary activity should be conducted, and what are the boundary objects (i.e., guiding questions, data chart of sunlight), as well the purposes of the post boundary activities (i.e., requiring doing higher cognitive levels of activities). And they will conduct learning in formal spaces in more structural way.

Table 1. The lesson exemplar of boundary activities

<table>
<thead>
<tr>
<th>Inquiry phase</th>
<th>Content</th>
<th>Learning spaces</th>
<th>Boundary Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Questioning and Context</td>
<td>Solar power is a clean energy. It is one of important resources for human being to obtain the electricity. Working with you group members to answer the following question: Which is the best time/time slot for us to make use of the solar power in daytime?</td>
<td>Classroom: Teacher assigns tasks and introduces the tasks with details in the classroom.</td>
<td>Pre-boundary activity</td>
</tr>
<tr>
<td>(WISE)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Sense-It Exploration</td>
<td>Using Sense-it to collect data of sun light and upload the data chart and share the evidence with classmates. The students are also required to comment on each other’s work.</td>
<td>Outside and home: Students collect data outsides and upload the data chart via Sense-it and review their classmates’ work and comment their work at home.</td>
<td>Boundary activities</td>
</tr>
<tr>
<td>(nQuire-it platform)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Sharing and discussion</td>
<td>Teacher presents all students’ work uploaded in nQuire-it platform, and identifies the quality of work and discuss with the students.</td>
<td>Classroom: The discussion and interaction focusing on the boundary objects in classroom.</td>
<td>Post-boundary activities</td>
</tr>
<tr>
<td>(WISE and nQuire-it platform)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Summary and reflection</td>
<td>The students summarize the science phenomena of sunlight by responding to the guiding question.</td>
<td>Classroom or home: The summary and reflection enable students to elaborate on their understanding of solar energy.</td>
<td>Post-boundary activities</td>
</tr>
<tr>
<td>(WISE)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusion and Future Research

In the paper, the conception of boundary activity is further coined in the field of mobile learning. The key elements of the boundary activity have been refined. The boundary object has been identified. The pedagogical principle of the boundary activities is incorporated into the design of mobile-technology supported science learning activity. Hence, the research will be a deeper attempt on the integration of curriculum development, mobile learning and learning in informal spaces in science. This will inform the science curriculum design and development supported by mobile technology.

Hofstein and Rosenfeld (1996) contended that “it would be useful if science educators would consciously utilize a wide range of out-of-school environments which foster science learning”. They preferred to adopt the “hybrid” view (rather than the dichotomy) that informal learning experiences can occur in formal learning environments as well as informal learning environments. They recommended that future research in science education should focus on how to effectively blend informal and formal learning experiences to significantly enhance the learning of science. The proposal of learning design based on boundary activity is the appropriate direction for addressing this.
In the future research, we will focus on unfolding the transformative process of students’ scientific concepts, inquiry skills and collaborative learning skills impacted by the boundary activity, snapshotting the reciprocal interaction between formal learning and learning in the informal spaces, finally changing teacher and students’ perspectives of the learning in the informal spaces and to enhance teacher competency on conducting learning activities in the informal spaces.

Acknowledgements

The paper is a part of work from the project: Boundary Interaction: Developing a Science Curriculum to Integrate Learning in the Informal Space (Project No. R3784), funded by Committee on Research and Development of The Education University of Hong Kong.

References


Patrick, P., C. Mathews, et al. (2013). Using a field trip inventory to determine if listening to elementary school students’ conversations, while on a zoo field trip, enhances preservice teachers’ abilities to plan zoo field trips. *International Journal of Science Education, 35*(15), 2645-2669.


Responsive eBook based on the Principles of Educational Interfaces

Ana Emilia OLIVEIRA*, Elza MONIER, Dilson RABELO, Fátima GATINHO, Aldrea RABELO & Marcelo Henrique Monier ALVES Junior

* anaemilia.unasus@gmail.com

Abstract: The growing popularity of mobile devices with internet access has created a unique opportunity for exploration in e-learning modality. When developing educational content for mobile devices, it is necessary to pay close attention to the technical aspects involved in educational interfaces. This paper presents a mobile educational object developed by UNA-SUS/ UFMA: the online responsive book. Its interface design is based on the Theory of Multimedia Learning and on the development standards for Responsive Web Design.

Keywords: Online eBook, responsive web design, educational interfaces.

1. Introduction

The internet is increasingly available to more people, as well as it is accessible by a wider range of devices, such as mobile phones and tablets. This fact has motivated the exploration of educational content to mobile devices, in order to increase the reach of the target audience.

Both the teacher and the designer are architects of the learning environments in the project and development of digital objects with educational interfaces. They should be aware that a good design of interface has to assure the student's attention and focus on the content, otherwise, the disorientation occurs, which causes the students to worry first about the particularities of the interface and later to the contents to be learned, according to Lima & Capitão (2003).

Therefore, when developing a multi-platform educational object adapted for mobile devices it is necessary to think of pedagogical principles to guide the interface design. It should be consistent with the purpose of the educational object, regardless of the access device.

The objective of this paper is to present a multimedia educational book adaptable to mobile devices through responsive web design technology. It is a learning object elaborated by UFMA through UNA-SUS (the Open University of Brazilian National Health System) to be used in e-learning courses.

2. Online eBook as an Educational Object in e-Learning

Technological evolution has made it possible to explore digital books, and some e-learning courses use this type of resource on a large scale. According to Vasileiou & Rowley (2008), digital books are educational objects containing texts and other elements, integrating the concept of a traditional paperback book with resources from the electronic environment. These resources are: search tools, crossed references, hyperlinks, multimedia objects and interactive tools.

In UNA-SUS’ experience with e-learning, the digital books, or online books, are adopted as the main source of presentation of contents of the institution's courses. They are books made available in a web accessible format.

The online book is one among many other resources gathered in the Virtual Learning Environment (AVA), which is an application used to manage institutions online courses. The book presented in this work was implemented to be used in any Internet browser that supports the use of HTML5. The informational design was organized with well-defined elements, in order to facilitate the navigation and visualization of content. Some Interface criteria were applied in its development. They are described by Reategui, Boff & Finco (2010):
• Use of images: according to the principle of multiple representation, it is better to represent an explanation with texts and illustrations, than just texts;
• Presentation texts: it is important to present the texts properly, noting some aspects such as contrasts between the sources and the background wallpaper, facilitating the reading of the texts;
• Orientation and navigation: to allow the users to locate in the resource what was done and what is available, among others;
• Interactivity: the user must interact with the learning object, being able to have a range of possibilities during its handling;
• Aesthetics: the learning object must have features that make it pleasant in its visualization;
• Affectivity: to allow the resources presented in the learning object to express affective states, such as moving images.

3. Responsive Online Book Interface

Responsive Web Design is a web interface development methodology that uses custom languages to construct pages in such a way that they respond dynamically to changes related to browser screen size. This concept originated from the publication of an article by Ethan Marcotte (2011), which shows a set of techniques that guarantee responsiveness to a web design.

By adopting a responsive design in an educational object, it is necessary to think about the processes of adapting the elements of its interface to various devices so that the pedagogical principles of educational interfaces are maintained in the transition between different screen resolutions. The responsive design can avoid inconsistencies in the fulfillment of the objective of the educational resource, respecting what was planned by the various actors in the process of teaching learning.

Next, we will show how the elements of the responsive online book interface were conceived and implemented, based on the technical principles of educational interfaces discussed in section 2.

3.1 Orientation and Navigation, Presentation of Texts and Visual Resources

Below we analyze the orientation and navigation scheme, text presentation and visual resources of the online book in two views: on a desktop and on a smartphone (figure 1).

Figure 1: Desktop (left) and smartphone (right).
The production of interfaces for different platforms has the challenge of bringing a clear orientation perception to all devices (Barbosa 2009). Analyzing the figure 1, we see that the menu icons are composed of images without text that intuitively suggest their functions, to properly use the space in various resolutions without losing meaning.

The navigation buttons clearly show the current page, how many pages there are, and how to move between pages. The header layout automatically adapts to the screen resolution change, maintaining the identification of the activity currently being developed in the course.

Fonts have an appropriate size whatever the screen size, but there is also an accessibility button to increase or decrease font size to the student's preference.

3.2 Fluid Images

In the construction of interoperable interfaces it is important that the images have adequate visualization in all the platforms since they are part of the educational content (Barbosa, 2009).

The images in the book are essential to illustrate the concept presented, making it more familiar and attractive to the student. Therefore it is important that the images continue to accompany the text independent of the screen resolution. To guarantee this, the images had to become fluid, changing sizes according to the access device.

3.3 Interaction and Layout

Regarding the design of multi-platform educational materials, to adjust interaction features is essential, mainly because they can be used through different devices with different settings (Barbosa, 2009). A special case is the change of the interaction that occurs in desktops and in smartphones. The interaction through touchscreen requires a touch area with a size suitable for the fingers of a human being. Therefore, when resizing icons that represent some function of the educational object, a minimum size limit must be established, to allow easy access to the features offered when touching the screen.

4. Conclusions

The main point of this work was the presentation of a responsive online book as a mobile educational object, developed by UNA-SUS / UFMA. The important features in the implementation of this online book for mobile devices were presented with the intention of contributing with other researchers and developers of responsive web design. The proposed mobile web interface adapts content in a way that does not harm contents visualization regardless of the screen resolution of mobile devices.

References

Encouraging System for Teaching Assistants to Advise Students during Programming Exercises

Yuuki YOKOYAMA & Hironori EGI
The University of Electro-Communications, Tokyo Japan
*y1730120@edu.cc.uec.ac.jp

Abstract: Computer programming courses form a basic field of study for both students who major in Computer Sciences and those studying a wide range of other sciences. However, such courses present difficulties for certain students. To overcome these issues, many universities have introduced teaching assistant (TA) programs to support students’ learning. However, TAs may not be able to effectively identify students having difficulties in class. Consequently, we designed a support system that enables TAs to better understand students’ learning performance during programming classes. The system employs head-mounted displays, which provide TAs with a direct visual representation of students’ current and historical class performances, enabling them to determine which students (if any) are facing difficulty and provide them with extra support.

Keywords: Teaching Assistants, Programming Exercises, Learning Logs, Advising

1. Introduction

Programming courses form a basic field of study for students majoring in Computer Sciences as well as those studying a wide range of other sciences. However, such courses present difficulties for certain students who may require extra support. This problem is compounded by the fact that typically, only one tutor is responsible for supervising a large number of students.

To improve this situation, many universities have introduced teaching assistants (TAs), whereby postgraduates are hired to help individual students in class. However, TAs can find it difficult to identify students who are experiencing issues. Ultimately, this can lead to students giving up and eventually failing.

To overcome this, we designed a support system that enables TAs to better understand the student learning process. The system collects performance data from the course server as well as visual feeds transmitted to the head-mounted displays (HMDs) worn by TAs, to help them determine which students are having difficulties with the class material and offer them learning support.

2. Programming Exercises and TA

Many universities run computer programming courses to teach students the basics of IT. Such courses are beneficial for students as they help them gain a better understanding of software design and structure. Classes mainly consist of students solving a series of exercises in a computer laboratory, and being taught by a single tutor. However, quite often, students experience difficulties in mastering such complex skills, especially if their background falls outside of Computer Science.

TAs play an important role in supporting students in programming classes. They have two basic tasks: (i) answering student queries about syntax or errors and (ii) identifying students who are experiencing difficulty and advising them on solutions. TAs are sourced from the student cohort and their position is closer to those of the students they are helping than as a tutor. However, in large classes, TAs can find it difficult to identify students requiring support, and therefore, improved ways of helping TAs to determine this are required.
3. Related Work

The most effective ways of teaching computer programming are important considerations within Computer Science courses. Students learn not only the grammar of programming languages but also coding techniques and problem-solving (Robins, 2003). Students are expected to work on programming exercises and practical assignments, on which tutors and TAs give feedback afterward.

Automated assessment, which provides feedback to students, is used in Massive Open Online Courses (Staubitz, 2015); however, this technique makes it difficult to monitor individual student’s learning progress.

The programming processes employed by the students represent the main way in which their learning progress is measured. To this end, a system that offers real-time, automatic monitoring, assessment, and evaluation of the code written by students as well as the production process employed by them was designed (Robinson, 2016).

Survey-based research has demonstrated that, according to TAs, verbal feedback is more effective than written feedback sent through an online system or e-mail (Rodgers, 2014). Thus, we propose a system that enables TAs to advise students based on their individual performance and learning needs.

4. Design

4.1 Requirements

The requirements for this classroom-based TA learning support system are as follows. TAs may not be able to determine which students are facing difficulties with the class material as it is difficult to monitor several learners in a large computer laboratory. The programming courses include complex elements and clear understanding requires considerable intellectual resources. The difficulties in mastering each of these elements vary among students. Therefore, by using our proposed system, TAs can identify specific elements that are not understood by particular students.

4.2 System Architecture

The courses are managed by the Learning Management System (LMS). The registered students are identified by their user IDs. Study materials are prepared and uploaded on the LMS. Achievement tests, assignments, and questions and comment about each lesson are submitted by the students. Attendance and absence are automatically checked by associating the students’ user IDs with the IP addresses of the PCs in the laboratory. Our proposed system collects attendance results, achievement test scores, and assignment grades from the LMS and provides it as a visual feed to the TA via the HMD.

4.3 Implementation

Student performance data are shown via visual feeds using the HMDs worn by TAs. Figure 1 shows a photograph of a TA wearing an HMD. The optical HMD (Vuzix M100) was chosen to allow unrestricted action and movement as TAs walk around the classroom supporting students. The proposed HMD system enables more convenient communication between TAs and students than tablets.
Figure 2 shows an overview and the procedures of the system. The three steps of the procedure are as follows.

4.3.1 Obtaining User ID of the Targeted Student

Printed QR codes associated with the IP address of each computer are provided on the back of each PC. Figure 3 shows the printed QR codes in a computer laboratory. The IP address of each computer corresponds to the user ID of the student accessing the LMS server. TAs move around the classroom during the programming class, and as the need to check a student’s learning progress arises, the TA simply scans the printed QR code associated with the student. The system takes the user ID of the target student corresponding to the QR code and sends it to the server.
4.3.2 Obtaining Three Types of Data

When the server receives the user ID of a particular student, it returns three types of data to the system for each lesson: (i) Attendance, (ii) Achievement Tests, and (iii) Assignments.

4.3.3 Visually Representing the Three Types of Data on the HMD

Figure 4 shows the visual representation of the three types of data—(i) Attendance, (ii) Achievement Tests, and (iii) Assignments—in a graph that appears on the HMD. Each block consists of the results from every lesson up to the present one with the student’s learning progress represented by green, yellow, or red color. For instance, in Attendance, green signifies attendance, red is absence, and yellow is late. In the Achievement Tests, green represents high scores and red represents low scores. The colors of the boxes in Assignments are green for submitted and red for unsubmitted. Empty boxes indicate that no data are available; for instance, no tests or assignments have been given yet.

![Figure 4. Visual representation of a student’s class-performance data](image)

In addition to the graphs, we designed a Seat Map that displays the class’s seating arrangements together with each student’s scores given in a separate graph. The Seat Map consists of two types of seating charts. The first displays the student’s scores from the current lesson, while the second shows their historical scores from past lessons. Figure 5 shows an illustration of a Seat Map for students in a computer laboratory. The server calculates student’s average scores in advance—which are represented by different colored boxes, where red signifies scores within the top 10% (subtracted from the average score), blue for scores within 10% of the top score (summed to the average and above), and green for scores between these two cohorts (red and blue). Seat Map can be accessed when TAs wish to monitor overall class learning progress.

![Figure 5. Seat Map showing student’s performance in the computer laboratory](image)
5. Experiment

This system was introduced to programming classes in order to evaluate its effectiveness in enabling TAs to identify students requiring support during lessons. The details of the programming classes are shown in Table 1. The system was introduced to four classes. The TAs wore HMDs and accessed student performance graphs during lessons #11 to #15 in Class A, lesson #13 in class C, and lesson #15 in Classes D and E. The number of instances where students had queries for TAs and received support were counted by observation. The lesson condition (i.e., whether HMDs were used or not) in each class is shown in Table 2. The page views (PVs) of student’s data and Seat Map data were calculated from the server logs. To provide a control condition, the number of requests for TA support made by students in the other classes where the HMD system was not used (i.e., B and F–K) was also measured.

Table 1: Detail of the lecture.

<table>
<thead>
<tr>
<th>Course title</th>
<th>Introduction to C Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td># of teachers</td>
<td>1</td>
</tr>
<tr>
<td># of TAs</td>
<td>2</td>
</tr>
<tr>
<td># of students</td>
<td>70</td>
</tr>
</tbody>
</table>

Length and frequency of lessons 90 min - once a week

6. Result

The incidences of questions asked by students (# of Qs) and the consequent advice given by TAs (# of As) during the lessons are shown in Table 3. No student questions or TA advice were observed in classes F–K. The results suggest that the impact of the HMD-based system on programming classes in terms of identifying students requiring support is unclear. It was evident that TAs’ behaviors varied between classes in term of offering student support to those in difficulty as certain TAs moved around the classroom and advised students freely, while others tended to sit in front of their PCs. Thus, one advantage of introducing the HMD-based system is that it encourages TAs to move around the classroom and offer support as they are required to scan the QR codes assigned to the students.

Table 2: Experimental conditions of classes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Lesson #11</th>
<th>Lesson #12</th>
<th>Lesson #13</th>
<th>Lesson #14</th>
<th>Lesson #15</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>HMD</td>
<td>HMD</td>
<td>HMD</td>
<td>HMD</td>
<td>HMD</td>
</tr>
<tr>
<td>B</td>
<td>Video</td>
<td>Video</td>
<td>Video</td>
<td>Video</td>
<td>Video</td>
</tr>
<tr>
<td>C</td>
<td>Video</td>
<td>N/A</td>
<td>HMD</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>D</td>
<td>N/A</td>
<td>Video</td>
<td>N/A</td>
<td>N/A</td>
<td>HMD</td>
</tr>
<tr>
<td>E</td>
<td>N/A</td>
<td>Video</td>
<td>N/A</td>
<td>N/A</td>
<td>HMD</td>
</tr>
</tbody>
</table>

Table 3: Number of questions and instances of advice during lessons.

<table>
<thead>
<tr>
<th>Class</th>
<th>Lesson #11 # of Q</th>
<th># of A</th>
<th>Lesson #12 # of Q</th>
<th># of A</th>
<th>Lesson #13 # of Q</th>
<th># of A</th>
<th>Lesson #14 # of Q</th>
<th># of A</th>
<th>Lesson #15 # of Q</th>
<th># of A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>0</td>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>D</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>N/A</td>
<td>N/A</td>
<td>4</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

The number of PVs during lessons #11 to #15 in class A are shown in Table 4, and those during lesson #13 of classes A and C and lesson #15 of classes A and E are shown in Table 5. PVs from lesson #15 (class D) were not collected due to server issues.

The analysis results showed that QR codes were scanned more often by TAs in the first lesson than in the following lessons, perhaps because of the necessity of getting used to the system. In addition, the Seat Map data were viewed by TAs more often than the graphs of student performance data.
In interviews with TAs, the visual representation of student performance data used in this HMD-based system was judged to be a suitable method for identifying students in need of extra support. However, TAs did report difficulties in talking to students even after they became aware of the learner’s need for support. The reasons for this require further research.

Table 4: Result of page view (PV).

<table>
<thead>
<tr>
<th>Lesson</th>
<th>#11</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV of student’s data (total)</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PV of student’s data (maximum number of a student)</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Seat Map PV of current lesson</td>
<td>31</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Seat Map PV of past lessons</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5: Result of page view (PV) in Lesson #13 and #15.

<table>
<thead>
<tr>
<th>Class</th>
<th>Lesson #13</th>
<th>Lesson #15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>PV of student’s data (total)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>PV of student’s data (maximum number of a student)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Seat Map PV of current lesson</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Seat Map PV of past lessons</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>

7. Conclusion and Future Work

In summary, we designed an HMD-based system that enables TAs to gain a clear overview of students’ learning progress during computer programming classes. The system collects student performance data from the course server and produces a visual representation of this for use by TAs. This method was introduced to real programming classes, allowing TAs to gain an overview of students’ learning progress and performance. However, the effectiveness of this system in terms of allowing TAs to more easily identify students requiring extra support with tasks was unclear. In addition, some TAs reported that wearing the HMD throughout the lessons was uncomfortable. In light of these findings, more research is required into increasing the efficacy of HMD-based systems in identifying students in need of extra learning support. Furthermore, it would be useful to compare the effectiveness of HMD-based systems with more traditional tablet-based approaches to identifying students in need of support during a class. It appears that TAs can benefit from being trained on effective monitoring of students in order to improve their ability to detect learners requiring enhanced support within the class environment.

Acknowledgements

This work has been partly supported by the Grants-in-Aid for Scientific Research (No. 15K01025) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan

References

Improving Primary Students’ Problem Solving Skills in Science Learning in a Seamless Learning Environment

Yanjie SONG\textsuperscript{a*} & Ka Man LUNG\textsuperscript{b}
\textsuperscript{a}Department of Mathematics and Information Technology, The Education University of Hong Kong
\textsuperscript{b}Man Kiu Association Primary School, Hong Kong
*ysong@eduhk.hk

Abstract: The paper reports on an empirical study aiming at improving students’ problem solving skills in science learning on “Plant Adaptations” in a seamless learning environment in a primary school. Two Grade 6 classes with 28 and 26 students respectively participated in this study. One class (G1) adopted project-based learning approach using a productive failure (PF) instructional design; the other class (G2) adopted project-based learning approach without using (PF) design. In G1, the students explored ways to grow plants in different conditions on their own without learning the related concepts and facilitated by the teacher about the misconceptions and assembly of conceptual knowledge after students’ completed their project; in G2, the students learned the related concepts first, then did the project. The learning activities spanned spaces across home, farm and school supported by mobile devices. Data collection includes pre- and post-domain tests, pre- and post-questionnaire on self-reported problem-solving skills, and student interviews. Both quantitative and qualitative methods were employed in the data analysis. The research findings show that compared to G2, the students in G1 improved domain knowledge after the project, and male students became more confident and critical in solving problems than female students, but female students developed better personal control than male students.

Keywords: Science learning, problem solving, conceptual understanding, productive failure, seamless learning

1. Introduction

In the digital age, problem solving is one of the 21st Century skills that we believe critical for preparing students in a global economy and society with increasing diversity, rapid change and efficient communication (Voogt & Roblin, 2012). Cultivating students’ problem solving skills means to engage students in solving a problem using different strategies, from multiple perspectives and with diverse modalities. This is in line with what is advocated in science education in Hong Kong that comprises a core component of the primary school General Studies’ curriculum. Doing science is to use the methods and procedures of science to investigate a phenomenon, test and develop understanding, solve problems and follow interests (Hodson, 2014). It promotes creativity through problem-solving process in authentic learning environments supported by digital technologies (The Education Bureau, 2011). However, in practice, science learning, in many cases, is still largely disconnected from learners’ daily life and confined to textbook learning (Anastopoulou et al. 2012). Students are passive knowledge receivers resulted from a lack of opportunities to be question askers, inquiry method designers and action-takers. How to develop students’ problem solving skills adopting innovative instructional design in a seamless learning environment is the main concern of this proposed project.

2. Literature
2.1 Productive Failure-based Instructional Design

Productive failure is defined as “a learning design that affords students opportunities to generate solutions to a novel problem that targets a concept they have not learned yet, followed by consolidation and knowledge assembly where they learn the targeted concept” (Kapur, 2015, p. 52). Productive failure instructional design first engages students in unguided problem solving to elicit their prior knowledge, particularly in the failure to solve the problem, followed by using this information to consolidate and aggregate new knowledge (Kapur, 2016). The failure stems from the fact that learners are commonly unable to generate or discover the correct solution to the novel problem by themselves; on the other hand, they are able to generate sub-optimal or even incorrect solutions to the problem, the process can be productive in preparing them to learn better from the subsequent instruction that follows (Kapur, 2014). Indeed, in science learning, generating “wrong answers” may help to focus students’ attention on the complexities and frustrations of a good investigation plan or design (Hodson, 2014).

2.2 Project-based Learning in a Seamless Learning Environment

Project-based learning is premised on constructivism and experiential learning and situated learning theories (Hmelo-Silver, Duncan, & Chinn, 2007; Kolb, A. Y., Kolb, D. A., Passarelli, & Sharma, 2014). In science learning, this approach aims to involve students in working at real-world problems in small groups and striving for solution options where the teacher acts as a facilitator (Brundiers & Wiek, 2013). The research literature shows that project-based learning can help students enhance learning performance in knowledge advancement and skill development, and motivate them to learn (Mioduser & Betzer, 2013). To achieve better learning outcomes, it is suggested taking into account the factors such as the complexity of the project, level of the learners, the learners’ prior knowledge and appropriate support (Thomas, 2000). In the digital age, learning becomes ubiquitous and seamless. Seamless learning refers to learning supported by mobile technologies across difference learning spaces in reconstructed contexts such as physical and virtual, formal and informal and individual and social spaces (Wong & Looi, 2011). However, how to develop students’ problem solving skills in a seamless learning environment still needs to be explored.

Thus this study addressed two research questions: (1) Was project-based learning using productive failure instructional design effective in developing students’ problem solving skills in a seamless learning environment? (2) Did students advance their science knowledge after this project-based learning in a seamless learning environment?

3. Methods

3.1 Research Design

This study was conducted in two Grade 6 classes (G1 and G2) with 28 and 26 students respectively on “Plant Adaptations” in a seamless learning environment in a primary school. There are four groups of students in each class. Each group had an iPad borrowed from the school. G1 adopted project-based learning approach using a “productive failure (PF)” instructional design; G2 adopted project-based learning approach without using (PF) design. In G1, the students explored ways to grow plants in different conditions on their own without learning the related concepts and facilitated by the teacher about the misconceptions and assembly of conceptual knowledge after students’ completed their project; in G2, the students learned the related concepts first, then did the project. Both G1 and G2 students investigated the problems related to plants and their environments in different spaces across home, farm, school and online supported by mobile devices in groups. The study lasted for two weeks.

The objectives of the project “Plant Adaptations” for both G1 and G2 are to choose two kinds of Rhizome plants to grow, and find out factors that influence the growth of the plants they choose. By doing so, students needed to work in groups to make their own plans to raise the plants. In order to understand better the factors that contribute to the growth of the plant, they usually prepared two or three plants of the same kind to raise in different conditions (e.g., different intensity of light, and amount water), observe, document and explain their process of growth.
3.2 Data Collection and Analysis

Data collection includes student pre- and post-domain tests, pre- and post-questionnaire on self-reported problem-solving skills, and student interviews. Pre- and post-domain tests had 8 multiple choice questions related to factors related to plant growth. The questionnaire was constructed based on the validated “Personal Problem-Solving Inventory” (Heppner & Petersen, 1982) in 3 dimensions: problem-solving confidence, approach avoidance style and personal control. It consisted of 23 questions on a 5-point Likert scale from 1 (never or rarely) to 5 (always). We received the valid responses from all the students in G1 (25 students) and G2 (26 students). The focus group interview had 5 questions. Two groups of students with 3-4 students in G1 and G2 respectively participated in the interview. The interview lasted about 35 minutes for each group.

Both qualitative and quantitative methods were employed in the data analysis. Students’ domain knowledge scores were rated according to the number of items that they selected answers indicating a traditional mindset on this domain knowledge. In total, the highest score is 8. Students’ problem solving scores were rated with a range from 1 to 5, on three sub-dimensions including problem solving confidence, approach avoidance orientation (negative rated), and personal control skills. Students’ focus group interviews were analysed using content analysis to do coding under the framework of the three dimensions of the questionnaire.

4. Research Results and Discussions

4.1 Pre- and Post-Domain Tests

First, students’ mean scores in the pre-domain test and the post-domain test were calculated in two ways: first they were calculated according gender; secondly they were calculated according to groups of G1 and G2. The results are shown in Table 1. ANOVA analysis results showed gender differences were insignificant in both pre-domain and post-domain mean test scores (F=0.017(1,52), p>0.05; F=1.135(1,52), p>0.05); group differences were significant in pre-domain mean test, while was insignificant in the post-domain test mean score (F=6.468(1,52), p<0.05; F=0.023(1,52), p>0.05). These results indicated that there was a difference in students’ domain knowledge of these two groups at the beginning of the course with G2 performed better than G1. However, the difference was diminished at the end of the project. This indicates that G1 improved their learning after the project.

Secondly, paired T-test analysis of students’ domain test scores on these two tests achieved the following results: there was significant difference in students’ pre-domain test scores and post-domain test scores in G1 (t (27)=2.777, p<0.01), but not in G2 (t(25)=0.000, p>0.05). In addition, there was significant difference in female students’ pre-domain test scores and post-domain test scores (t (28)=2.736, p<0.05), but not male students (t (24)=1.000, p>0.05). The results indicate that in G1, students’ traditional mindset on the domain’s knowledge transformed significantly after the project, and females benefited more than males. By triangulating participants’ statements about reasons of transformation, it can be concluded that the project helped them to build up scientific domain knowledge.

Table 1: A comparison of pre-domain and post-domain test scores

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>3.68</td>
<td>1.228</td>
<td>3.03</td>
<td>1.149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>25</td>
<td>3.64</td>
<td>1.578</td>
<td>3.44</td>
<td>1.635</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Group:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>3.19</td>
<td>1.327</td>
<td>3.19</td>
<td>1.575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>4.11</td>
<td>1.315</td>
<td>3.25</td>
<td>1.236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>3.67</td>
<td>1.39</td>
<td>3.22</td>
<td>1.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2 Students’ Problem Solving Skills

Students’ mean scores of problem solving skills measured before and after the course were calculated, and shown in Table 2. Paired T-test about the difference between G1 and G2 were conducted. The results of this analysis show that no significant differences in G1 and G2 were found in students’ problem solving skills on all three dimensions of problem-solving confidence, approach avoidance style and personal control before and after the project. However, there was a significant increase in students’ problem solving confidence of the male group, but not the female group (p<0.01). The results indicate that male students benefit more than female in their problem solving confidence by doing the project using productive failure instructional design. In addition, a close to significant decrease in male students’ personal control during problem solving was also found (p=0.058) in the two groups. The results indicate that the course may be able to raise more personal control challenges for the male students.

Table 2: Scores of students’ problem solving skills achieved before and after the course

<table>
<thead>
<tr>
<th>Gender</th>
<th>Confidence</th>
<th>Approach avoidance</th>
<th>Personal control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Before</td>
</tr>
<tr>
<td></td>
<td>Mean-SD</td>
<td>Mean-SD</td>
<td>Mean-SD</td>
</tr>
<tr>
<td>Group1</td>
<td>3.02-0.541</td>
<td>3.69-0.607</td>
<td>3.19-0.531</td>
</tr>
<tr>
<td>Group2</td>
<td>3.12-0.617</td>
<td>3.23-0.555</td>
<td>3.19-0.584</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.00-0.700</td>
<td>3.19-0.426</td>
<td>3.34-0.516</td>
</tr>
<tr>
<td></td>
<td>2.96-0.676</td>
<td>2.96-0.434</td>
<td>3.27-0.579</td>
</tr>
<tr>
<td>Total</td>
<td>3.10-0.558</td>
<td>3.16-0.505</td>
<td>3.16-0.505</td>
</tr>
</tbody>
</table>

We interviewed 2 groups with 3-4 members in G1 and G2 respectively. We categorize the themes in three problem solving dimensions of problem-solving confidence, approach avoidance style and personal control. The results show different views of students in the project-based learning supported by mobile technologies in G1 and G2. For example, regarding problem-solving confidence, when asked how they knew that the plant’s growth is relevant to the intensity of light, the students in G1 reported that they learned it by their own observation, capturing the photos, and making comparison and contrast with those growing indoors. Their new findings were: the leaves of the plants growing outdoors usually spread around; but the leaves of the plants growing indoors unusually went upwards for light; while students in G2 grew bean sprouts and found that the growth of green beans growing in water and soil was different. When asked what the differences lay, they said they did not know and did not have the intension to find it out later.

In terms of approach avoidance style, when asked how they made their plan to grow the plant in different conditions, the students in G1 reported that they searched information on their own, and discussed in groups via WhatsApp and face-to-face. Finally, they worked out a plan of growing the plant; while students in G2 responded that their teacher provided guidance for the plant growth (e.g., Don’t put too much water.).

Finally, regarding personal control, when asked whether they felt frustrated and grew a sense of failure in failing to grow the plant, the students in G1 responded that instead of having a sense of failure, they had a sense of success as they successfully grew other kinds of plants and learned more about the conditions or environments for specific kind of plants’ growth, and were glad to know more about the functions of plants; while students in G2 reported that they felt a bit disappointed about failing in growing the plant, but did not mention what they thought of the plants that they raised successfully.

Thus, from the examples we noted that the students in G1 were engaged in active learning by exploring and finding out solutions on their own without fearing of failures. The results were in line with the findings of studies using productive failure instructional design in mathematics learning (Kapur, 2014). However, students in G2 needed guidance during their inquiry and did not form the habit
of challenging the problems they could not understand. Mobile devices supported the students in documenting the growth of the plants they studied.

5. Conclusion

The paper reports on an empirical study to examine the effectiveness of G1 of the project-based learning using productive failure instructional design on students’ science learning and problem solving skills development in a seamless learning environment in a primary school compared with G2 without the instructional design. The research findings show that compared to G2, the students in G1 improved domain knowledge after the project, and male students became more confident and critical in solving problems than female students, but female students developed better personal control than male students. Mobile technology helped their project-based learning in documenting the growth of the plants with camera or recording functions, and communicating with group members. Future research needs to be done to examine the effectiveness of the productive failure instructional design in a seamless learning environment in schools at a larger scale.

Acknowledgements

The study was funded by The Education University of Hong Kong under Internal Research Grant 2015-16 [R3734].

References


Blockino: a Tool with an Emphasis on Educational Robotics Assisting the Teaching of Programming Logic

Higo SAMPAIO*, Mauro SILVA*, Aldrea RABELOb, Carla MARQUESc & Marcelo MONIERa

*IFMA – Instituto Federal do Maranhão, Brasil
bUFMA – Universidade Federal do Maranhão, Brasil
cUniversidade CEUMA, Brasil

Abstract: For entering students in Computer Science, in addition to having affinity with the area, it is highly necessary to acquire knowledge in Programming Logic. However, not all entering students in the technical or graduation courses find it easy to learn this subject. As an attempt to contribute to change this reality, the creation of a tool called Blockino has been developed. Blockino includes the concept of Educational Robotics, helping the teachers and professors in the process of construction of knowledge in programming logic through the use of a programming language in the blocks with robots.

Keywords: blockino, educational robotics, programming logic

1. Introduction

The market of technology demands for more qualified professionals every year, and unfortunately in Brazil there is a lack of qualified workforce to meet this demand. According to González (2013), the country has an average percentage of 87% of evasion in the courses of this area. This happens due to a weak mathematic and logic basis in the Brazilian elementary schools, leading to an ever harder effort from the universities. The Federal Institution of Education, Science and Technology of Maranhão (IFMA) offers technical courses and graduation courses in the area of computing science, and it is currently experiencing a high rate of school evasion. The professors from the Computer Science Academic Department have identified that among the main factors related to the school evasion is a major difficulty in the subject Programming Logic, which is the first contact of the entering student with computing science.

A proposal for the solution of this problem is to seek a more interactive learning process in the subject of Programming Logic. Technology, especially Educational Robotics, is a strong tool to go deeper in the learning process. According to Zilli (2004), this technology helps to develop a range of skills such as logical thinking, tests and hypothesis formulation, applying theories developed from specific activities, problem solving through analysing hits and misses, among others.

Accordingly, the present work aims to validate a tool called Blockino, that consists into a mobile application using visual programming based in blocks, in order to manipulate the robot through bluetooth technology, facilitating the first contact of the entering student with programming logic, in a more dynamic way. The tool has been used as an experiment with a group of entering students of the Technician Course of Computing Science at IFMA.

This article is structured in the following manner: the section 2 defines the proposed tool and how it was conceived; the section 3 presents the experiment carried with the tool; and the last section presents the concluding remarks and future intentions for the continuity of the tool.

2. Educational Robotics

The discussion regarding the Information and Communication Technologies (TICs) in the educational environment has been going on for years, and has as main goal to improve the teaching quality in the
learning environments. It is important to highlight that although there are so many ideas and proposals, there is a supremacy of tools designed exclusively for software development (Sampaio, Miranda Borges, 2010). However, Sampaio, Miranda and Borges (2010) state that many efforts are made to change this reality. Within the proposals, the insertion of robotic in the educational system is a very strong one, as it stimulates construction of knowledge inside the classrooms and promotes student’s motivation. This educational component provides a more ludic and challenging environment for the students, that are subjected to find solutions to specific problem situations.

The proposal to use robotics in educational environments was developed by the researcher from the MIT Computer Science and Artificial Intelligence Laboratory, Seymour Papert, who advocates the use of computer as a pedagogical tool. His defence is based in the constructivist theory by Piaget, that aims to explain human intelligence emphasizing the relationship of a person with the environment during the process of human cognitive development. Stem from this theory, Papert idealizes an adaptation called Constructionism, which is different from Jean Piaget’s Constructivism, and it is based in building knowledge through concrete experiences, stating that the learning process is built in parallel with previous knowledge designed in practical experiments (Zilli, 2004).

Educational Robotics (also known as Pedagogical Robotics), uses industrial robotics, as construction, automation and robot-devices control, in a learning environment, in order to teach the learning concepts. According to the Brazilian Education Interactive Dictionary (DIEB), the term Educational Robotic can be understood as a learning environment that embraces mechanic artifacts which are organized in a set to make a functional model controlled via software. Thus, programming is used to somehow facilitate a proper operation for the idealized model.

The University of Joensuu, in Finland, has a project called Children Club, which consists in an environment where children between 10 and 14 years old participate in experiments with pedagogical robots in laboratories, under the supervision of their mentors, who are students from the university. The assessment of this experiment showed that the use of pedagogical robotics, if compared to the traditional methodologies, promotes a more motivational, fun and dynamic learning, consisting in learning based in problem-solving, overcoming the traditional methods. This pedagogical robotics workshops are characterized by many different forms and objects, but in general, they hold a learning process in which the students are subjected to reading instructions and manuals to use reach their goals, or experiment spontaneous construction and robotic prototype programming (Eronen et al., 2002).

To Zilli (2004), the arrival of pedagogical robotics helps develop a range of skills such as logical thinking, tests and hypothesis formulation, applying theories developed from concrete activities, problem solving through analysing hits and misses, among others. Therefore, this methodology is available for the teachers, to facilitate the comprehension of more complex topics during expository lectures (Schons, Primaz, Wirth, 2004). The challenges created by the teachers and professors, allow the students to have collaborative and constructive discussions on the topics, engaging and motivating the students to develop practical solutions.

3. Blockino

The Blockino was idealized with objective to create a tool to assist the teaching of the school subject Programming Logic. It is composed by an Android application and using the Google’s Bockly Library, and a small robot made with Arduino electronic prototyping board, suffering direct effects through bluetooth technology.

Figure 1 shows a structure of communication between the elements (Student, Blockino and Robot), that represents all the architecture specifications used in the project, identifying basically three types of views: components, software and hardware.

The components’ view describes the scenario to use tool and a communication between the elements involved. So, it is the scenario that students use application in smartphone or tablet and use interface based in blocks, to send commands to handle the robot.

The software's view represents the interaction between the two applications presented in the project, the mobile application and robot embedded programming. The Blockino mobile application, represents the interface of the project user, it is there that student will manipulate blocks programmable to execute functions reflected in robot. The application had its layout structure based in the technologies HTML, CSS and Javascript, with adaptation and integration the Blockly tool to manipulate blocks.
After the interface user was finished, used the Apache Cordova tool to create applications in Android, then allowing to use mobile device with this operational system. Figure 2 shows the main screen of the mobile application.

**Figure 1.** Communications structure between the elements.

**Figure 2.** Main screen of the mobile application.

This screen represents the objective of the application. There are two bars with functionalities icons, one in top, other in the left side and one little garbage in the under right side to discard the blocks that will not be used.

The sidebar contains icons to create the blocks: control blocks, arithmetics, logicals, variables and functions. Among them, only blocks of functions were custom to call the possibilities of robot movement, and all others were used the Google's Blockly tool. Blocks like “walk_forward”, “walk_backward”, “move_right”, “move_left”, “led_up”, “led_off” and “calculate_distance”, are commands that represent movement of robot, and can be grouped to intensify the movement complexity. An example to do the robot walking non stop, can be resolved with a kind of control block - represent the repeat structure - grouped to block “walk_forward”.

On top of the bar there are icons with general functionalities on application: create code, help and execution. The button “generate code”, allows the visualization of code written in text format, with objective to be the alternative to view code. The button “help” shows the application's iconography, and the button “execution” sends commands grouped by bluetooth.

Beyond mobile application, it is necessary an embedded code on the robot, that has as main function to receive and translate the commands sent by the application. Figure 3 shows this translation of commands by programming embedded on robot.
The embedded application is guided exclusively by the loop function that is requested continuously while robot is on in the energy. This function has the objective to read space reserved by bluetooth technology to save the information received. When the loop function finds any string, it slices the received string into small strings according to the delimiter "$", defined previously with function to concatenate more commands in only execution of programmable blocks. These strings separated are analysed by other function, that has the goal of test and connect the string with name of the responsible methods to handle the robot. For example, this function could identify the command “walk_forward”, making the robot move forward.

The way the physical components of project interact define the hardware's vision. At this stage, the user uses a device with operation system Android 2.3 or superior, and a bluetooth connection to send the code. The robot was made with pieces and components compatible to Arduino platform and consist in a vehicle with three wheels, in which two back rears are connected directly to the motors and the front wheel counts for support and guidance. In its structure, there are to supporting basis overlaid to for a better attachment and organization of the other electronic components.

4. Experiment

4.1 Experiment Application

For the development of the Blockino application, researches were made on how to use educational robotics in classroom and extract significant results from the experiment. Therefore, in order to assess the project, the application of a practical-theoretical experiment was applied to students in the second year of IFMA’s Computing Science technical high school. These students were invited to participate in the experiment having previous programming knowledge.

In order to encourage collaborative work, four students took part in this experiment which lasted for two hours in two stages: theoretical class and practical class. The experiment started with a brief theoretical explanation regarding the selection and repetition structures which was fundamental to review some concepts and prepare the students for the challenges. In the practical class a short introduction about the Blockino tool was given concerning points such as main objective, browsing characteristics and application usage. The introduction to the tool was made necessary for making it possible the challenges resolution.
In total, five proposed challenges happened as practical activities, accounting: two selection structures, two repetition structures and one final challenge with two themes mixed. Below are three examples of the challenges applied:

a) To create a simple selection structure in which its true logic expression makes the robot walk forward and turn on a LED;

b) To create a for repetition structure which makes the robot walk forward and turn left five times starting from value zero for the interaction variable;

c) To create a for repetition structure which makes the robot walk forward and turn right five times starting from value two and test if the value is equal to four after each interaction. If the result is true the robot must turn on the LED and walk backward.

At the end of the experiment the students answered an exercise for consolidation of acquired knowledge, filling a questionnaire about usability and acceptability aspects. Below there are examples of the questions used in the exercise:

a) Is the tool visually attractive?
b) Did the tool allow the completion of the desired tasks?
c) Is the tool easy to use?
d) Did the tool enable a fast resolution of the exercises?
e) Are the icons used on the tool familiar?
f) Did you manage to assimilate the presented subjects about programming logic?
g) Can the tool be considered a support for the teaching of Programming Logic?

4.2 Experiment Analysis

The main objective on creating a support tool for the teaching of programming logic was positively considered at the end of the experiment. The teachers involved agreed that the tool could become an alternative resource, making the classes more dynamic, interactive and interesting. Figure 4 represents a graphic generated according to the resolution of the questionnaire applied to students at the end of the experiment.

5. Final Considerations

With the development of the work it was possible to conclude that, seeking alternative teaching techniques to try to increase the reach of learning is a good solution especially for students who demonstrate difficulty in absorbing content. The Blockino tool contemplates the fundamentals of Pedagogical Robotics, which uses robots to streamline the teaching process. The level of interaction of this one becomes greater because it is a tangible experience, and thus has greater chances to arouse more
interest on the part of the students. This tool, which integrates elements of block programming and bluetooth communication, has proven to be a viable alternative to assist teachers in the teaching of programming logic, contributing to avoid demotivation and consequently failure rates.

Currently, the possibility of migrating from the mobile application to a web service is being discussed, thus allowing parallel access of users to try to solve the problems presented by the teachers in a more competitive and fun way. The gamification elements can be thought of as an auxiliary tool in student engagement.

As future work, we intend to cover this experiment in a larger group of students and who do not have prior knowledge in programming, including also other fields of education such as undergraduate.

References


Effects of Prior Knowledge of High Achievers on Use of e-Book Highlights and Annotations

Misato OI*, Fumiya OKUBO, Yuta TANIGUCHI, Masanori YAMADA, Shin’ichi KONOMI
Faculty of arts and science, Kyushu University, Japan
*oimisato@gmail.com

Abstract: To identify “good performance,” this study analyzed the highlighting and annotating action logs of undergraduates during their e-book usage. To reveal “good performance,” the study focused on the learning behavior of high achieving students. Few highlights and annotations were observed for both rich knowledge and poor knowledge high achievers. Moreover, in the spontaneous usage of e-books outside the classroom, high and poor knowledge students did not display differences in highlights and annotations.

Keywords: e-book, preview, review, annotation, highlight

1. Introduction

In the past decade, the popularity of big data and analytics has increased (Picciano, 2012), including in education. Many researchers have highlighted the potential of big data in education (e.g., Daniel, 2015; Long & Siemens, 2011; Picciano, 2012). For example, “big data can provide institutions of higher education the predictive tools they need to improve learning outcomes for individual students as well ways ensuring academic programmes are of high-quality standards” (Daniel, 2015).

To improve learning and teaching, Kyushu University introduced a single platform learning system (Mitsuba, or M2B) that was based on a common learning management system (Moodle), an e-portfolio system (Mahara), and an e-book system (BookLooper). The e-book system enables students to browse e-book materials before/during/after lectures, anywhere and anytime, using their personal computer (PC) or smartphone. All user actions performed on this e-book system, such as page flips and opening a material, are recorded as learning logs and automatically sent to the University’s database when a network connection is available. The e-book system provides the additional functions of bookmarking, highlighting, annotating, and searching. By the end of April 2017, approximately 45,000,000 logs (Moodle: 28,000,000; Mahara: 1,000,000; BookLooper: 16,000,000) were collected from approximately 20,000 students. Analysis of educational big data from M2B provided insights into the activities of students, such as browsing patterns with respect to quiz scores (e.g., Shimada, Okubo, & Ogata, 2016), effective learning behavior (e.g., Oi, Okubo, Shimada, Yin, & Ogata, 2015a; Oi, Yin, Okubo, Shimada, Kojima, Yamada, & Ogata, 2015b; Oi, Yamada, Okubo, Shimada, & Ogata, 2017a; Oi, Yamada, Okubo, Shimada, & Ogata, 2017b, Yamada, Shimada, Okubo, Oi, Kojima, & Ogata, in press), and predictive modeling (e.g., Okubo, Shimada, Yin, & Ogata, 2015).

To develop an effective feedback and/or intervention system, the concept of “good performance” needs to be clarified to reveal to students their goal (Sadler, 1989). The aim of the present study is to identify “good performance” by analyzing e-book logs of M2B. In previous studies (Oi et al., 2015a, b; Oi et al., 2017a, b), we focused on learning behavior, namely, covering the same content before (preview) and after (review) its learning in a class session. Undergraduates’ performance of such preview and review was analyzed based on e-book logs categorized as follows: if a log was recorded before a class session in which the same e-book was used as a textbook, it was a preview log, and if after, a review log. The main findings are as follows: (1) preview is more deeply related with academic achievements than review (Oi et al., 2015a; 2017b), (2) relatively low achievers attempted to perform previews, but they tended to give up easily on the endeavor (Oi et al., 2017b).

As a first step to identifying “good performance,” the present study analyzed the highlighting and annotating action logs of undergraduates during e-book usage. To understand new information, annotating and highlighting are useful techniques, and these are assumed as a valuable part of the
process of learning (Glover, Xu, & Hardker, 2007). A study reported that university students commonly used annotation to identify key parts of the document (Ovsiannikov, Arbib, & McNeill, 1999). As pointed out by a classical study (Ausubel, 1960), if a student has prior knowledge of the contents of a course, it may help the student’s learning by acting as an advance organizer. In other words, students’ prior knowledge of contents of a course could affect their learning behavior, and that “good performance” may differ according to the amount of knowledge processed. For example, a student who has less knowledge has to verify technical words and the relationships between fundamental concepts, particularly while performing a preview. However, a student who has considerable knowledge does not need to verify such words and relationships. Based on this hypothesis, we examined differences between rich vs. poor knowledge students and their preview vs. review performances. To reveal “good performance,” we mainly focused on the learning behavior of high achieving students.

2. Methods

2.1 Participants

E-book logs were collected from 110 undergraduates enrolled in an information science course (from April 4 to July 26, 2016, 14 sessions); these logs were also analyzed in Oi et al. (2017a, b). The objective of the course was for students to understand the fundamentals of information and communication technology (ICT). For the assessment of students’ prior knowledge of ICT, the students took a placement test before beginning the first session; the test comprised questions from the Information Technology Engineers Examination (https://www.jitec.ipa.go.jp/index-e.html). Students also took a midterm and an end-term examination during the 8th and 14th sessions, respectively. The teacher did not provide clear instructions to the students regarding the use of highlights or annotations.

After completing all of the sessions in the course, students were given their final score, which was converted into a grade (i.e., A: 90–100, B: 80–89, C: 70–79, D: 60–69, and F: less than 60). The final scores were calculated for each student from his/her mid-term examination score (30%), end-term examination score (30%), short reports (10%), and attendance (20%).

For analyses, we excluded logs from students who did not take the placement test (n = 4), the mid-term examination (n = 4), or the end-term examination (n = 2); those who did not submit any short report; and those who received a grade “F” (n = 1). We considered the scores of the placement test to represent the level of students’ prior knowledge of ICT (i.e., the contents of the course). In categorizing students with rich or poor knowledge, they were divided into four groups according to the quartile of the scores of the placement test.

Table 1 summarizes the number of students according to a combination of the quartile of the placement test and the grade.

<table>
<thead>
<tr>
<th>Placement test</th>
<th>Final grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>8</td>
<td>15</td>
<td>5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>7</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>24</td>
<td>50</td>
<td>16</td>
<td>9</td>
<td>99</td>
</tr>
</tbody>
</table>

2.2 E-book Logs

The total number of e-book logs was 447,650. Table 2 presents a sample of the e-book logs.
### Table 2: Sample of e-book logs.

<table>
<thead>
<tr>
<th>logid</th>
<th>userid</th>
<th>operationname</th>
<th>operationdate</th>
<th>contentsid</th>
<th>deviceid</th>
<th>memo</th>
<th>page_no</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTG7z</td>
<td>4RjqBr</td>
<td>OPEN</td>
<td>2015/3/31 14:57</td>
<td>012ABC</td>
<td>76UjvV</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GbycT</td>
<td>4RjqBr</td>
<td>PREV</td>
<td>2015/3/31 14:57</td>
<td>012ABC</td>
<td>76UjvV</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>My0bl</td>
<td>4RjqBr</td>
<td>PREV</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>qUQxf</td>
<td>4RjqBr</td>
<td>NEXT</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1JCv7</td>
<td>4RjqBr</td>
<td>ZOOM</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>zN3Gl</td>
<td>4RjqBr</td>
<td>ZOOM</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GLJPt</td>
<td>4RjqBr</td>
<td>ZOOM</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AHLfX</td>
<td>4RjqBr</td>
<td>ADD MARKER</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>nMKVx</td>
<td>4RjqBr</td>
<td>ADD MEMO</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td>ギガ</td>
</tr>
<tr>
<td>94xjjM</td>
<td>4RjqBr</td>
<td>PORTRAIT</td>
<td>2015/4/2 10:21</td>
<td>012ABC</td>
<td>UFQq7C</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

### 3. Results and Discussion

#### 3.1 Number of e-Book Logs

First, we confirmed whether the number of e-book logs varied vis-à-vis students’ final grade and rank in the placement test. Figure 1 presents the average number of e-book logs for each group. To account for groups with a small $N$ (e.g., $n = 1$ for A-C [Placement-Final grade] and D-D), SDs are not shown. To examine the differences, we performed one-way ANOVAs on the number of e-book logs with groups of placement test and final grade, respectively, as between-subject factors. As the $N$s of some groups were considerably small, we did not employ a two-way ANOVA. The ANOVAs revealed a significant difference among the final grade, $F(3, 95) = 11.42, p < .0001, \eta^2 = 0.27$. However, the difference among the groups of the placement test was not significant, $F < 1$. Figure 2 presents the average number of e-book logs for each final grade group. Multiple comparisons on difference among the final grades with Bonferroni correction (corrected $p < .05$) revealed that the number of e-book logs of group A was significantly higher than that of the other three groups. No other significant differences were observed between the groups. These results indicate that (1) high achievers use e-books more frequently than middle and low achievers, and (2) prior knowledge of the course did not significantly affect the number of e-book logs.

![Figure 1. Average number of e-book logs for each group.](image-url)
3.2 Highlights and Annotations

We focused on high achievers with a final grade of A. To examine whether the level of prior knowledge affects high achievers’ annotations and highlights, we analyzed the frequency of highlights and annotations of the rich knowledge group (i.e., A-A) and poor knowledge group (i.e., D-A) (see Table 1) during preview, class session, and review. Table 3 presents the number of highlights and annotations for each student. Only two students from the rich knowledge (A-A) group and one from the poor knowledge (D-A) group used highlights during the spontaneous e-book usage outside the classroom (i.e., both preview and review). In other words, the use of highlights by both rich and poor knowledge high achievers was relatively minimal. Although DA01 added 42 highlights, we could not determine whether this log indicated a characteristic of poor knowledge high achievers or simply that of the individual student.

Both groups showed few annotations throughout their preview and review. Only AA03 showed annotation logs during preview; however, the text fields of the annotations were blank. In other words, AA03 clicked the annotation button but did not write anything. Moreover, during the review and class session, six students clicked on the annotation button but no text was written. We checked the annotation command logs of all students: 209 of the 588 logs contained text. For example, an annotation text of one student is “Merge sort is three times faster.” (This annotation text was translated from Japanese). The remaining 379 logs did not have text.

Table 3: Frequency of highlights and annotations during preview and review.

<table>
<thead>
<tr>
<th>Highlight</th>
<th>AA01</th>
<th>AA02</th>
<th>AA03</th>
<th>AA04</th>
<th>AA05</th>
<th>DA01</th>
<th>DA02</th>
<th>DA03</th>
<th>DA04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preview</td>
<td>15</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Review</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation</th>
<th>AA01</th>
<th>AA02</th>
<th>AA03</th>
<th>AA04</th>
<th>AA05</th>
<th>DA01</th>
<th>DA02</th>
<th>DA03</th>
<th>DA04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preview</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Review</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

High achievers did not write annotation text, and a third of them used the highlight function in preview. To confirm whether the use of few annotations and highlights is a characteristic of high achievers, we summarized the number of students who used these functions for each grade throughout all of the logs. Table 4 summarizes the number of students who used highlights or annotations. Almost
half of the students did not use these functions, and thus, we could not directly compare the number of annotations and highlights with respect to the final grades. We performed Chi square tests on the ratio of the students who used the functions and those who did not. The Chi square tests revealed the absence of significant differences among the grade groups for both the highlights, $\chi^2(3) = 6.463, p = .091$, and the annotations, $\chi^2(3) = 6.151, p = .104$. Thus, there was no significant difference in the grades of the ratio of students who used the functions of highlighting and annotating.

An implication of the above results is that without explicit instructions to use highlights and annotations, students may not actively use these functions of the e-book system. However, these analyses were performed based on the e-book logs; records of usage of other electronic devices and paper notes were not included. More enriched functions (e.g., free sketching and automatic summarization of annotations and highlights) could perhaps encourage students to use annotation and highlights and thus enable deeper analyses of learning behavior.

Table 4: Number of students who used highlights and annotations.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highlight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16</td>
<td>32</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>18</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Annotation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16</td>
<td>27</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>23</td>
<td>11</td>
<td>6</td>
</tr>
</tbody>
</table>

4. Conclusion

As a first step to identify “good performance,” the present study analyzed the highlighting and annotating action logs of undergraduates during e-book usage. To determine “good performance,” we mainly focused on the learning behavior of high achieving students. Few highlights and annotations were observed for both rich knowledge and poor knowledge high achievers. In the spontaneous usage of e-books outside the classroom, high and poor knowledge students did not exhibit differences in their use of highlights and annotations.

Acknowledgements

These research results have been achieved under the theme of “Research and Development on Fundamental and Utilization Technologies for Social Big Data” (178A03), as Commissioned Research for the National Institute of Information and Communications Technology (NICT), Japan.

References


Promoting Extrinsic Motivation Based on Result of LMS Quiz

Yasutaka ASAI* & Hironori EGI
Graduate School of Informatics and Engineering,
The University of Electro-Communications, Japan
*a1730003@edu.cc.uec.ac.jp

Abstract: When learning alone using LMS, students can find it difficult to maintain motivation because self-managed learning ability is required. In this research, we focused on the role of praise as an extrinsic form of motivation because it is thought that it is necessary to support learning. We designed a function of praise giving for students based on their results in a quiz with LMS which was implemented in quiz module of Moodle. It was introduced to the programing exercises and experiments were conducted to test the results.

Keywords: Extrinsic Motivation, Praise, Learning Log, LMS, Quiz

1. Introduction

Learning Management Systems (LMS) are web-based applications specialized for browsing learning materials, taking quiz, submitting reports, and notification of grades. LMS has been widely introduced to educational institutions. Students are expected to proceed with self-managed learning without restrictions of time and place with using the system. However, self-managed learning tends to cause students to experience difficulties with maintaining motivation as they are required to plan their own study schedule and carry this out by themselves.

In this research, we focus on effectiveness of praise, an extrinsic form of motivation. The praise function is designed to encourage students to study based on the result of an LMS quiz. It is evaluated within the context of an actual programming course in order to investigate how praise affects the learning motivation of each student.

2. Motivation and Learning

2.1 Extrinsic and Intrinsic Motivation

Motivation for learning can be considered from two perspectives; (i) intrinsic, such as framing learning activities as fun, and (ii) extrinsic, such as framing tasks as necessary in order to graduate. Learning activities without intrinsic motivation are thought to hamper maintain motivation for learning over the long term.

According to Keller (2009), a feedback loop of motivation in learning activities is often found. Confidence due to the sense of accomplishment when the activity is completed and curiosity about the next activity due to this sense of satisfaction awakens intrinsic motivation for learning. Stimulating learning activities is essential to improve this motivation feedback loop. Although extrinsic motivation has less influence of self-managed learning on students than intrinsic motivation, it is expected to encourage them and raise their intrinsic motivation. Introducing extrinsic motivation is considered to be one of the effective ways of stimulating the learning process.

2.2 Motivation by Praise

Hurlock (1925) conducted an experiment on children for 4 days concerned with how praise affected their test scores. It revealed that the scores of children who were given praise showed consistent
improvement and therefore, praise seems to be effective for stimulating learning as a form of extrinsic motivation.

Shimada (2012) conducted an experiment that praised students using a web-based learning system. A function of praising and scolding was implemented to maintain learning motivation of these students. The result revealed that the students who were given praise were more motivated in performing the learning tasks. This cohort was more pleased when they answered questions correctly and appropriate messages were shown. However, this cohort’s scores were almost the same as students who studied without praise. The reason was that the tests were perhaps too easy for the students. However, as this experiment was conducted only once, its reliability is debatable. Such kinds of praise should be introduced to actual courses on several occasions in order to gain a more reliable indication of its effectiveness.

3. Promoting Motivation with LMS

The key idea of this research is giving praise based on learning activities. This maintains and improves their learning motivation as a using result of a quiz taken via the LMS. It is assumed that praising learning activities especially supports less motivated students. Hurlock (1925) reveals that scolding is not effective for boosting learning motivation so scolding was not implemented.

Scores and repetition of the quiz ensured that the impression that the messages were generated at random was reduced.

The praise function was implemented via the quiz module of Moodle. It was introduced to a programming course in a university as an experiment. Changes in student’s learning were investigated in order to evaluate the effective praise in improving motivation.

4. Praising Function

4.1 Design of Praise Message

The praise messages were shown to students individually after finishing the quiz. The flow indicating the praise messages displayed is shown in Figure 1. An example of a praise message is shown in Figure 2.

![Figure 1. Flow of indicating praise messages](image-url)
For example, if the student’s score was high, an admiring message such as "That’s good" or "Wonderful" was displayed. On the contrary, if the score was bad, encouraging messages such as "You challenged!" or "Let's review a little" were displayed. In addition, if the student took the quiz repeatedly, messages such as "You’re really improving" "You’re getting better" were displayed. The details of choosing messages are described in the next section.

The function records chosen praise messages, quiz scores, the number of time the quiz is taken and the quiz ID for each student. Eighty kinds of praise messages were prepared and displayed one after another so that the same message was not repeatedly displayed.

4.2 Choosing Method of Praise Messages

In order to choose the proper praise message, "Score Rate" and "Time of Taking Quiz" were used. Score Rate is the percentage of the score. Time of Taking Quiz is the amount of time taken to conclude the quiz. The results of the quiz were classified into five types of praising plans considering these parameters. The classified plans are defined as praise plans A-E respectively. The rules and thresholds for determining the praise plan are shown in Figure 3.
The rules and thresholds for determining the praise plan were set up with reference to the result of the quiz in the previous year. Table 1 shows the criteria assumed for the praise plans.

Table 1. Criteria assumed for the praise plans

<table>
<thead>
<tr>
<th>Praise plan</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Score Rate is 100%</td>
</tr>
<tr>
<td>B</td>
<td>Many examinations and high score rate</td>
</tr>
<tr>
<td>C</td>
<td>Few examinations and high score rate</td>
</tr>
<tr>
<td>D</td>
<td>Many examinations and low score rate</td>
</tr>
<tr>
<td>E</td>
<td>Few examination and low score rate</td>
</tr>
</tbody>
</table>

The quizzes were classified into four categories defined as: "Knowledge / exercise,” "Coding / exercise,” "Knowledge / review” and "Coding / review.” The labels of categories are shown in Table 2. Different rules and thresholds of the praise plan were set for the knowledge quiz and coding quiz.

Table 2. Labels of categories

<table>
<thead>
<tr>
<th>Categories</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>Quiz on choosing an option or filling in blank</td>
</tr>
<tr>
<td>Coding</td>
<td>Quiz on operating the terminal or coding</td>
</tr>
<tr>
<td>Exercise</td>
<td>Quiz about the contents explained in the current lesson</td>
</tr>
<tr>
<td>Review</td>
<td>Quiz about the contents explained in a previous lesson</td>
</tr>
</tbody>
</table>

5. Experiment

We conducted an experiment in C language programming courses held in one semester. The courses are for first-year students in a university of science and engineering. The courses consist of fifteen lessons once a week. Students used the praise function from the beginning of the 11th lesson to the end of the 15th lesson. Each lesson is taught in computer laboratories. 373 students in six classes took 6629 quizzes during the experiment. After completion of the experiment, a questionnaire on attitudes to the quiz was filled out by the students both in introduced classes and ordinary classes for comparison. The questions are shown in Table 4.

6. Results

6.1 Effectiveness of Praise Function

We compared how many times the quiz was taken between the two cohorts; the class with the praise function and the ordinary class. Figure 4 shows a comparative graph of the times the quiz was taken per student. The shaded part in Figure 4 is the period in which the praise function was introduced. Looking at the graph, classes that introduced the praise function showed an increase in the number of times the quiz was taken. There is a possibility that students took the quiz more times because of the novelty of the praise function. Considering the prior period of the experiment in Figure 4, the students in the praise condition might be intrinsically well-motivated to take quiz more times than the ordinary condition subjects. There are many elements that affect student’s motivation and it is uncertain that the praise function is the specific cause which increased the times the quiz was taken unconditionally. However, it is possible that the praise function influenced the learning effectiveness of the students.
6.2 Evaluation of the Praise Plans

Table 3 shows the number of displayed messages about the knowledge review quiz obtained from the experiment. The columns show the results of classified messages into praise plans A-E and the number of messages. The rows show the messages displayed in each quiz of lessons #11, #13, #14 and every lesson of the experiment. The results of lessons #12 and the other categories are included in the total row.

Table 3. Number of displayed messages in each lesson and praise plan

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>#11</td>
<td>205 (38%)</td>
<td>7</td>
<td>105 (19%)</td>
<td>46 (9%)</td>
<td>180 (33%)</td>
<td>543</td>
</tr>
<tr>
<td>#13</td>
<td>108 (20%)</td>
<td>40</td>
<td>84 (15%)</td>
<td>79 (14%)</td>
<td>244 (44%)</td>
<td>555</td>
</tr>
<tr>
<td>#14</td>
<td>115 (21%)</td>
<td>29</td>
<td>78 (14%)</td>
<td>66 (12%)</td>
<td>264 (48%)</td>
<td>552</td>
</tr>
<tr>
<td>Total</td>
<td>2095 (32%)</td>
<td>229</td>
<td>1149 (17%)</td>
<td>504 (8%)</td>
<td>2652 (40%)</td>
<td>6629</td>
</tr>
</tbody>
</table>

As the total number of displayed messages in each praise plan, the number of praise plan A and E is larger than the other plans, while the number of praise plans B and D is fewer than the other plans. The reason for this imbalance in the displayed messages is thought to be caused by the inadequate rules and thresholds for determining praise plans.

The highest rates of correct answers are calculated as best scores for each student per quiz. The average of the best scores from all the quizzes is 72.5%. The average number of time the quiz was taken is 1.62 times. In addition, the rate of taking the quiz (which is taken only once for each student) is 59%. The average number of times the quiz was taken before reaching a perfect score was 1.9 times, thus, it was inadequate to set the threshold of taking the quiz to three times or more.

The praise plans A-E of lessons #13 and #14 seems to be more balanced than those of lesson #11 and every other lesson of the experiment. The reason is perhaps that quizzes in these lessons are considered to be very difficult. The quiz in lesson #13 deals with the topics of pointers and the quiz in lesson #14 deals with the topics of character strings using pointers. From the results of taking quiz in each lesson, the rules and thresholds for determining the praise plan should be designed relating to the quiz scores.
6.3 Result of Questionnaire

As a result of the questionnaire, Table 4 shows the average of the scores divided into the praise condition and the no-praise condition.

Table 4. The average scores of questionnaire

<table>
<thead>
<tr>
<th>#</th>
<th>Questions</th>
<th>Praise</th>
<th>No Praise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Was the quiz on knowledge difficult?</td>
<td>3.55</td>
<td>3.69</td>
</tr>
<tr>
<td>Q2</td>
<td>Was the quiz on knowledge long?</td>
<td>3.45</td>
<td>3.38</td>
</tr>
<tr>
<td>Q3</td>
<td>Was the quiz on operation difficult?</td>
<td>3.98</td>
<td>3.94</td>
</tr>
<tr>
<td>Q4</td>
<td>Was the quiz on operation long?</td>
<td>3.51</td>
<td>3.72</td>
</tr>
<tr>
<td>Q5</td>
<td>Did you master the knowledge needed with the quiz?</td>
<td>3.61</td>
<td>3.58</td>
</tr>
<tr>
<td>Q6</td>
<td>Did you master operation with the quiz?</td>
<td>3.51</td>
<td>3.43</td>
</tr>
<tr>
<td>Q7</td>
<td>Did you try until you understood the quiz fully?</td>
<td>3.31</td>
<td>3.29</td>
</tr>
<tr>
<td>Q8</td>
<td>Have you confirmed the results of the quiz?</td>
<td>3.67</td>
<td>4.08</td>
</tr>
</tbody>
</table>

Looking at the average scores of the questionnaires, there was no difference between the praised class and the no-praise class. As consequence of this, there are many elements other than LMS that constitute the lesson; it is possible that those elements are not dependent on LMS-affected scores. For example, the teaching method was different for each teacher in charge of teaching and the tendency of motivation for learning and the learning activities carried out by each class was different. This could have had a huge influence. In addition, the questionnaire design may not have been adequate to investigate the influence of praise on motivation.

7. Conclusion

In this research, we developed an experiment to maintain and improve motivation for learning using praise based on student’s learning activities with LMS and conducted an experiment on its effectiveness. In the Moodle quiz, we developed means of classifying students using their score rate and the number of times they took the test and displayed a praise message image accordingly. In addition, this was introduced to university lecture and an experiment was carried out. Based on the results, further analysis was undertaken.

In terms of future work, we think that set the threshold value based on the data of the last year should be considered in more detail and that the threshold should be altered with reference to the log data in order to make it possible to set the threshold for deciding on a praise plan according to the difficulty level of the quiz. This research does not investigate the content of praise messages. It is necessary to conduct a questionnaire about the content of the message and to investigate which type of messages are most effective for boosting motivation for learning.

References


An Electronic ID System Using a Smart Phone

Jaewook KIM*, Takashi TACHINO & Yasuhiito KISHI

*Department of Media and Culture, Shoin University, Japan
*jwkim@shoin-u.ac.jp

Abstract: This paper proposes a personal identification method using a smart phone that displays a face image, character data and a two-dimensional including personal identification data. We have constructed the prototype personal identification system using a smart phone and a Web server to identify each student who attends a class. The effectiveness and the security of the proposed personal identification method are investigated from a wide viewpoint that is not limited to the prototype.

Keywords: smart phone, personal identification, two-dimensional symbol, security

1. Introduction

Personal identification is important in various information systems. We have already proposed the authentication method using a face image for a Web server. This paper proposes a personal identification method using a smart phone that displays a two-dimensional including identification data. Although plastic cards and smart cards are used as ID (identification) cards, the risk to lose these cards are large. In some countries including Japan, many people have color display smart phones with data communication functions. If a smart phone is used for personal identification, it is possible to dynamically display all of the identification data on the color display of the smart phone. Therefore, this is an electronic ID system.

Two-dimensional symbols are constructed by means of extending the functions of barcodes. Two-dimensional symbols have characteristics such as large recording capacity and high-level error correction capability using a Reed-Solomon code. There are various kinds of two-dimensional symbols. Two-dimensional symbols can be classified as a stack type and a matrix type. The stack type symbol such as PDF417 is constructed by stacking multiple low-height barcodes. The matrix type symbol such as QR Code is constructed to store a black or white pattern called as a smart that corresponds to a pixel of an image. By adding the two-dimensional symbol in the personal identification system using a smart phone and a Web server, the two-dimensional symbol including identification data can be quickly read by the CCD camera attached to a computer and decoded to the text data. The security of the personal identification data is strengthened. Because it is possible to apply encryption and digital signatures to the two-dimensional symbol.

The current student identification card in a college or a university is mainly used for the personal identification at the term examination, and at the time when each student receives a certificate from the office of the school. The problems of the current student identification card are that the face photograph size is not large enough to precisely distinguish each student, and it takes several hours or more to receive the reissued identification card when a student forgets to bring the identification card or loses it. We have constructed the prototype personal identification system using a smart phone and a Web server to identify each student who attends a class. The proposed method can be extended to the general personal identification method using a smart phone.

2. Fundamental Concept of the Proposed Personal Identification Method

The fundamental concept of the proposed personal identification method using a smart phone is summarized as follows.

(1) The personal identification data are stored in the database of a Web server (or a computer center).
(2) Each user receives the personal identification data from the Web server on each smartphone. The timing when the user receives the personal identification data depends on each application. It is possible for the user to request the personal identification data of the Web server. It is also possible that the Web server automatically sends the personal identification data at the appropriate time decided by the application.

(3) The smartphone displays the personal identification data such as character data (name, address, etc.), image data (face image, etc.). The checking person checks these displayed data.

(4) The two-dimensional symbol including the personal identification data is additionally displayed on the smartphone. The purpose is to read the displayed data by using a CCD camera with the two-dimensional symbol decoder, and automatically process the identification data at the terminal computer of the checking person.

(5) The two-dimensional symbol is secured by using encryption and authentication techniques [4]. Additionally, the two-dimensional symbol can include the timestamp to limit the valid time for the personal identification data.

3. Implementation of the Prototype

3.1 Overview of the Prototype System

To confirm the correctness of the proposed personal identification method using a smart phone, we have constructed the prototype personal identification system for a college based on the proposed method. The overview is shown in Fig. 1. At first, the administrator inputs the personal identification data for the Web server. The Web server creates a two-dimensional symbol including the personal identification data. The personal identification data including the two-dimensional symbol is transmitted to the student's smartphone at the pre-specified time. At the time of the beginning of a class, each student shows each smartphone displaying the two-dimensional symbol data to the CCD camera with a decoding function for the two-dimensional symbol. Therefore, the attendants are confirmed by using the personal identification data of the two-dimensional symbol.

![Figure 1. Overview of the prototype system.](image-url)
Next, the student name and the student number are recorded in the roll database and can be displayed on the manager’s notebook computer. The Web server is constructed by using Linux operating system, Web server software and Script language.

3.2 Smartphone of Students

An example of the displays of the electronic ID using a smart phone is shown in Fig. 2. The display of the smart phone can show the face image, the character data and the two-dimensional symbol. The two-dimensional symbol includes the identification data, and can include the digital signature generated by the Web server for the identification data. By using the two-dimensional symbol, the identification data can be quickly read by the CCD camera and the security of the identification data is strengthened. QR code is used for the two-dimensional symbol of the prototype system.

![Figure 2. An example of the displays of the electronic ID using a smart phone.](image)

3.3 Manager’s Terminal Device

The manager’s terminal device receives the decoded data of the two-dimensional symbol from the CCD camera with the two-dimensional symbol decoder that changes the content of the two-dimensional symbol to the text data, displays the identification data on the screen of the terminal, and displays the list of identified persons such as the student name and the student number.

4. Security Improvement for the Web Access by the manager’s portable device

Since the transmission of a face image is performed directly, the size of transmission data is large and there is the danger of leakage of the face image from the communication line. In order to resolve these problems, the following schemes are investigated.

4.1 Message Authentication for the Image

**Scheme-M:**

**Step M1:** The face image for authentication is stored in the manager’s portable device. At the time of logon, the manager’s portable device creates the message authentication code MAC-1 of the transmitter side by using the face image and the transmission time. Next, the message authentication code MAC-1 is transmitted from the Manager’s portable device to the Web server.

**Step M2:** The Web server receives the data from the terminal, and creates the message authentication code MAC-2 of the receiver side by using the registered face image in the Web server. Then, if the message authentication code MAC-1 of the transmission side is coincided with the message authentication code MAC-2 of the receiver side, and only if there is no inconsistency between the transmission time and the received time considering the communication line delay, the user's logon is
permitted. The outline of processing of Scheme-M using the message authentication codes of face images is shown in Fig. 3.

The message authentication code is a technique to detect the forgery of a message. The message authentication code is a message digest of data that is created by using a common key cipher algorithm.

The procedures to create and verify the message authentication code (MAC) using the face image are described in more detail as follows.

Step 0 (Setup): The manager’s portable device holds face image V1, encryption key K and the encryption program. The Web server holds face image V2, encryption key K and the encryption program using a common key cipher algorithm.

Step 1: The manager’s portable device creates message authentication code MAC-1 by encrypting face image V using key K, and transmits MAC-1 to the Web server.

(Note: It is unnecessary to transmit face image V1. This is different from the usual usage of a message authentication code that transmits both the MAC and the original message.)

Step 2: The Web server receives MAC-1. Next, the Web server creates message authentication code MAC-2 by encrypting face image V2 using key K. If MAC-1 is equal to MAC-2, the Web server decides that face image V1 is equal to face image V2.

5. The Security of the Electronic ID System using a Smartphone

The attacks and counter measures with regard to the electronic ID system are shown as follows. These investigations are based on a wide viewpoint and are not limited to the prototype system.

**Attack-1**: The unauthorized user who has a smart phone may access the fraudulent Web site created by him, without accessing the correct Web site, and he may display the fake identification data on the display of the smart phone.

**Counter measures against attack-1**: Attack-1 can be defended when the correct Web site attaches the digital signature in the electronic ID data such as a two-dimensional symbol.

**Attack-2**: The unauthorized user may display the two-dimensional symbol that is illegally obtained from the correct person on the smart phone.

**Counter measures against attack-2**: Since a face image is displayed on the smart phone, the check person can compare the face image to the live face. Additionally, since the two-dimensional symbol includes the timestamp to limit the valid time for the ID data, the two-dimensional symbol may become invalid when it is used illegally.

**Attack-3**: The intruder for the Web server may modify the personal identification data in the database of the Web server.
Counter measures against attack-3: It is very difficult to intrude the secure Web server, to modify the personal identification data, and to generate the two-dimensional symbol including the digital signature. The security measures of a Web server is beyond the scope of this paper (e.g., see [1]).

6. The Merits of the Personal Identification Method using A Smartphone

Table 1 shows the comparison of ID systems among the ordinary ID card system using a plastic card, the electronic ID system using a smart phone without the two-dimensional symbol, and the electronic ID system using a smart phone with the two-dimensional symbol. Comparison items are the risk for losing the ID card, precision of identification, the time for checking the identification data, and the security of the ID system.

Table 1: The comparison of ID systems.

<table>
<thead>
<tr>
<th>Items</th>
<th>The ordinary ID card system such as a plastic card</th>
<th>The electronic ID system using a smart phone without the two-dimensional symbol</th>
<th>The electronic ID system using a smart phone with the two-dimensional symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>To reduce the risk for losing the ID card</td>
<td>C</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>The precision of identification</td>
<td>C</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>The time for checking the identification data</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>The security of the ID system</td>
<td>C</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Notes: A(Excellent), B(Good), C(Unsatisfactory)

The proposed personal identification method using a smart phone has the following merits for the electronic ID system using a smartphone. These merits are not limited to the prototype system.

- As the personal identification data can be downloaded from the Web server anytime, it is unnecessary for each person to worry about losing the identification card. Since a smart phone device is frequently used for a phone call and an email besides the electronic ID using a smart phone, the possibility to lose the smart phone is smaller than the possibility to lose a plastic ID card or a smart card.
- As the face image displayed on the smart phone is large enough in comparison with the photograph size of the ordinary plastic ID card, the precision increases when the check person compares the displayed face image to the live face.
- As the personal identification data can be directly read by using the CCD camera with the decoder of the two-dimensional symbol, the data can be automatically processed by the check person’s terminal device such as a notebook computer. As the personal identification data can be downloaded from the Web server anytime, it is unnecessary for each person to worry about losing the identification card. Since a smart phone device is frequently used for a phone call and an email besides the electronic ID using a smart phone, the possibility to lose the smart phone is smaller than the possibility to lose a plastic ID card or a smart card.
- By adding the two-dimensional symbol for the electronic ID system using a smart phone, the forgery becomes more difficult and the time for checking the contents of the personal identification data is reduced.
7. Discussion

It is possible to use a specific image pattern instead of a two-dimensional symbol. However, it is difficult to devise a new image pattern, read it by a CCD camera, and quickly decode the specific image pattern to the text data. Therefore, two-dimensional symbols are better than a specific image pattern as for the electronic ID system.

The proposed personal identification method can be extended to the general-purpose electronic ID system using a smart phone, for example, the electronic ID system for a company or an organization, the electronic passport for international travelers, etc.

8. Conclusion

A personal identification method using a smart phone that displays a two-dimensional has been proposed. By using the proposed method, it is possible to construct an electronic ID (identification) system using a smart phone. By using a smart phone for personal identification, it is possible to display all of the identification information on the color display of the smart phone. A prototype personal identification system using a smart phone and a Web server has been constructed to identify each student who attends a class in a college. The display of the smart phone can present the face image, the personal character data and the two-dimensional symbol. The two-dimensional symbol includes parts of the personal data. By using the two-dimensional symbol, the personal identification data can be quickly read by the CCD camera and the security of the personal identification data is strengthened. The proposed system can be extended to the general-purpose ID system using a smart phone. The problem for further study is to construct the other applications using the idea of the proposed personal identification method.

References

Social Network Analysis of Teacher’s Role in Students’ Online Discussion Community

Chih-Ming CHU
Department of Computer Science and Information Engineering, National Ilan University, Taiwan
cmchu@niu.edu.tw

Abstract: Because of the popularization of mobile devices and social media platforms, an increasing number of teachers have begun to adopt online student discussion groups as a method of teaching. Some teachers have asked students to add teachers to their discussion groups, while others have made no such request. The present study recruited college students and teachers for a “special project” course. The online discussion group without the teacher was defined as the control group, while the group with the participation of the teacher was defined as the experimental group. Without knowing that it was an experiment, the students were asked to use the online group for discussion for a period of three months. The messages and interactions of the two groups were then analyzed. The results showed that the number of messages of the control group was significantly higher than that of the experimental group, and the experimental group was relatively passive in leaving messages and less active in interactions. Additionally, the key figure of the community changed from being a student (group leader) in the control group to the teacher in the experimental group.

Keywords: social network analysis, online discussion, network learning community, degree centrality

1. Introduction

With the prevalence and rapid development of e-learning platforms, more educational institutions have begun to use e-learning platforms as a supporting tool for teaching (Sun & Gao, 2017; Swaggerty & Broemmel, 2017)). However, although learners may use the e-learning platforms out of curiosity in the beginning, they tend to gradually stop using the platform because of a lack of patience and external stimulation over time, which has been found to lead to a negative impact on learning performance (Wilfried, Jantina, Sanne, & Geert, 2011). For that reason, some studies have utilized emails to send regular reminders to learners to keep them motivated (Hodges, 2008; Hodges & Kim, 2010), or utilized short message service messages to send learning materials to learners on a regular basis (Hayati, Jalilifar, & Mashhadi, 2013). Nevertheless, the above methods still require learners to study on their own instead of encouraging them to interact with other learners. Such a learning strategy may result in learners developing feelings of loneliness and alienation and lead to high absenteeism (Liu, Magjuka, Bonk, & Lee, 2007; Rovai, 2001; Mayer & Moreno, 2003). Online discussion communities have been found to effectively establish an online learning environment and stimulate students’ participation (Colachico, 2007; Conrad, 2005). In addition to improving learning performance, the interaction between community members may also reduce feelings of isolation (Rizzuto, 2017; Hramiak, 2010). Studies have pointed out that peers are very important in the learning process (Wei, Hung, Lee, & Chen, 2011). Bannan-Ritland’s (2002) research showed that interaction is crucial for online learning, as it constitutes a part of the teaching goals; therefore, increasing interaction is conducive to the implementation of online courses. From the perspective of educational theory, having learning companions has been found helpful in the acquisition of knowledge as well as beneficial to learners’ emotional states and degree of social interaction (Khlaif, Nadiruzzaman, & Kwon, 2017; Kim & Baylor, 2006). Exchanging information between peer learners may also facilitate learners’ acquisition of more knowledge (Lan, Sung, & Chang, 2007). When an individual studies on his/her own, the learning process is easily interrupted because of factors such as frustration, while support from peer learners tends to facilitate the continuity of the learning process (Dean, Harden-Thew, & Thomas, 2017;
Dawson, 2010). Studies have revealed that peer interaction and discussion are conducive to problem-solving and knowledge absorption (Chen, Chang, & Wang, 2008). Seeking help from learner peers through online discussions is one way of solving problems (Guan, Tsai, & Hwang, 2006). Through participation in interactive peer learning, students can develop problem-solving skills and new knowledge acquisition abilities (Chou & Tsai, 2002). Additionally, online interaction can provide support and counseling to students with learning difficulties (Chen, Chang, & Wang, 2008). However, most research on the impact of online discussions on community learning appears to focus on the interaction during the communication process (Dominguez-Flores, & Wang, 2011; Liu, Magjuka, Bonk, & Lee, 2007). Few studies have investigated learners’ roles in the community and the influence of teachers’ involvement in online discussion. Therefore, the present study explored learners’ roles in online discussion communities as well as the influence of teacher’s participation on learners’ performance in a discussion.

2. Literature Review

2.1 Online Learning Community

Communities that are formed around the core purpose of learning are defined as learning communities. This concept was first proposed by Alexander Meiklejohn (1932) and John Dewey (1933). They stated that a learning community allows learners to develop personal viewpoints and remain connected to the concurrent realities of the learning context; therefore, a learning community is an environment that allows a group of individuals to exchange necessary knowledge and information during learning processes and provides individualized learning (Kochtanek & Hein, 2000). An online learning community refers to a group of learners that expand and develop knowledge and abilities both as a unit and as individuals in an online interactive learning environment (Hanna, Glowacki-Dudka, & Conceicao-Runlee, 2000), which is similar to a virtual learning environment and includes information sharing, discussion, and file downloading functions (Barry & Asiedu, 2017; Gillespie, Boulton, Hramiak, & Williamson, 2007). Ke and Hoadley (2009) suggested that an online learning community is an organization that implements learning in a virtual and supportive environment. In sum, an online learning community can be defined as a group of individuals with common learning objectives who participate in learning activities and share knowledge in a virtual community, establish gradual trust during interaction, and thereby develop a close-knit community relationship.

2.2 Social Network and Social Network Analysis

The concept of social networks was first proposed by Barnes (1954) to demonstrate a group of real-life social relations and mainly to explore the relations between individuals and the impact of relational structure on those individuals. Schultz-Jones (2009) suggested that social networks refer to the links between an individual and other individuals, including close relationships, secondary relationships, and other relationships that facilitate the connection between individuals and other individuals, as well as events and objects. A social network contains three key factors: actors, relationships, and linkages (also known as ties) (Mitchell, 1971). Actors are considered by the nodes of a network; they refer to the people, events, and things that define the network, and therefore, they constitute the main body of a social network. An actor usually belongs to many different networks simultaneously and may play different roles in each network. Relationships include the “existence” of a relationship and the “type” of relationship. Actors interact with one another because of the existence of certain relations. Different types of relations and corresponding content lead to a multitude of network environments. Common relations include transaction relations, communication relations, instrumental relations, sentiment relations, authority/power relations, and kinship and descent relations. Next, when an actor intends to establish a relationship with other actors, he/she must build the relationship directly or indirectly through a given path. Such basic links between actors are known as “ties.” Based on the closeness, ties can be divided into “strong” and “weak.” By analyzing the strength of ties, subgroups and brokers within the network can be identified. Social network analysis is an analysis method developed based on social statisticians to study social structures, interpersonal relationships, organizational systems, and
group interactions (Apperson & Beckman, 1999). In social network analysis, nodes and links represent the relationships within a group; by considering nodes as group members and links as the relationships between members, one is able to clearly demonstrate the structure of the social network, the types of relationships between members, and how they influence one another. In the Dictionary of Psychology, Corsini (2002) defined community roles as a high-level concept of behavior that is composed of a series of behaviors exhibited by an individual (Corsini, 2002). However, within the academic field of social network analysis, there is a disagreement on the definition of “role.” Some studies claim that the notion of role depends on the notion of position. Position refers to a collection of actors that have similar social behavior, links, and interactions (in relation to the interaction of other actors) within the same relational network (Wasserman, & Faust, 1994). Therefore, a role is not defined by an actor’s own attributes but rather by the types of relations between actors and their positions within the network (Wasserman, & Faust, 1994). Other studies argue that position and role have an interdefining relationship. Position is generated by the interaction between actors. However, once a position is formed, it is able to shape the relationships between actors. Therefore, the role is the link between the actors and positions. The present study adopted the viewpoint of community role and determined learners’ role in the community. The present researchers then utilized social network analysis to explore the relations and influence between community members throughout the entire network structure and adopted the concept of centrality to examine members’ roles in the community. The center was used to measure an actor’s influence or power (Wasserman, Faust, 1994). The greater the centrality of an actor, the more influential he/she was defined to be.

3. Research Method

The present study recruited six juniors (four male and two female) from the Department of Information Engineering from a university in Taiwan as research subjects. The course utilized for the experiment was dubbed a “special project design.” Without being informed that the course was a research project, students were asked to use Facebook’s Messenger function to form two discussion groups: a control group and an experimental group. The control group was composed of six students, while the experimental group was composed of six students and their teacher for the course. After three months, social network analysis software Ucinet 6.506 and Netdraw 2.138 were employed to examine the differences in the number of messages exchanged and interactions between members between the two groups.

4. Results and Discussion

The number of messages exchanged by the control and experimental groups are shown in Table 1. S1, S2, S5, and S6 referred to the male students, and S3 and S4 were the female students. S3 was the group leader, and S7 was their teacher. After a three-month discussion within the Facebook Messenger groups, members of the control group exchanged 1,029 valid messages, accounting for 63.3% of the total number of valid messages of the two groups. Meanwhile, the experimental group exchanged 597 valid messages, accounting for 36.7% of the total number of valid messages of the two groups. This finding shows that, compared to members in the group with teacher involvement, members in the group with no teacher involvement tended to be more active in discussion participation. This situation is similar to the situation of students who feel too inhibited to raise their hands in class and express their viewpoints publicly, and yet willingly exchange their opinions privately with other students. Additionally, the number of messages sent by each student, in the order from large to small, was S3 > S5 > S2 > S6 > S4 > S1 in the control group and S2 > S3 > S1 > S5 > S4 > S6 in the experimental group. It can be seen, based on the changes in the order, that S1 was notably active in the experimental group with the presence of the teacher, suggesting that S1 was driven by the teacher’s presence to gain recognition. In terms of response rates, the order from high to low was S1 > S6 > S3 > S2 > S4 > S5 in the control group and S3 > S2 > S1 > S5 > S4 > S6 in the experimental group. Although S1’s response rate was the highest in the control group (73.1%), he only sent 26 messages, much lower than the number of messages sent by S3 (377 messages). It is worth noting that the number of messages and the
response rate of S4 were relatively low in both groups, suggesting that S4 was not active in the participation of the discussion groups. Moreover, compared to the control group, S6’s response rate in the experimental group decreased substantially, indicating that S6 was less active in the group with teacher involvement. In sum, S1, S4, and S6 appeared to require extra attention from other members and the teacher so as to improve the effectiveness of the group discussion.

Table 1: The number of messages exchanged by the control groups and experimental groups.

<table>
<thead>
<tr>
<th>Student code</th>
<th>Gender</th>
<th>Number of times to be responded</th>
<th>number of messages</th>
<th>response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>control groups</td>
<td>experimental groups</td>
<td>control groups</td>
</tr>
<tr>
<td>S1</td>
<td>M</td>
<td>19</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>S2</td>
<td>M</td>
<td>69</td>
<td>34</td>
<td>193</td>
</tr>
<tr>
<td>S3</td>
<td>F</td>
<td>166</td>
<td>32</td>
<td>377</td>
</tr>
<tr>
<td>S4</td>
<td>F</td>
<td>25</td>
<td>4</td>
<td>81</td>
</tr>
<tr>
<td>S5</td>
<td>M</td>
<td>48</td>
<td>6</td>
<td>262</td>
</tr>
<tr>
<td>S6</td>
<td>M</td>
<td>41</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>S7</td>
<td>M</td>
<td>431</td>
<td>463</td>
<td>1029</td>
</tr>
</tbody>
</table>

Figure 1 illustrates the interactions of the control group during the discussion. The lines represent the interactions between two members. The denser the lines were, the more were the messages exchanged between two members. It can be seen from Figure 1 that interaction was present between all members. Moreover, S3 appeared to have more interactions with other members and more frequent interaction with S5, showing that S3 played the central role in the community, which defined his identity as the team leader. A further comparison between Figure 1 and Table 1 revealed that although S1 sent the least messages in the group, his response rate was the highest and he interacted with all members. It is apparent that under the attention of group members, S1 was still able to maintain interaction with others rather than being marginalized. Figure 2 demonstrates the interactions of the experimental group during the discussion. It can be seen clearly that S7 (the teacher) was the central figure of the community; the messages sent by all members reduced significantly, forming an S7-centered discussion environment. With the exception of S4 and S6 (no interaction was found between them), the majority of the members (including the teacher) had interactions with all other members; however, the interactions between S7 and other members appeared to be more dominant. Further, S7 had more interaction with S3 (team leader). It was apparent that the discussion among the members of the experimental group mainly centered around the teacher; interactions between students were limited.
5. Conclusions and Suggestions

Many teachers in the teaching will require students to group the way to set up community online discussion. Some teachers have asked students to add teachers to their discussion groups, while others have made no such request. The following conclusions and suggestions are given in this study. 1. If there is no teacher in the student's online discussion group. Although students can take the initiative to discuss and the number of messages are more. But the quality of discussion not grasp. There is no way for the teacher to discuss the process of the students. Only to see the final result presented. 2. If the teacher joins the online discussion group and plays an active role in leading the discussion topic. The performance of the students will be more passive and silent. 3. It is recommended that teachers should join the student's online discussion group, should more observation and less intervention. If necessary, make comments for students. Avoid intervening discussions and too many speeches.

References


Exploring the Nature of Teacher’s Ongoing Feedback to Pupil using iPad

Weiyun CHEN\textsuperscript{a,b}, Yaofeng XUE\textsuperscript{a,b} & Xiaoqing GU\textsuperscript{a,b}\*  
\textsuperscript{a}Department of Education Information Technology, Faculty of Education, East China Normal University, Shanghai, China  
\textsuperscript{b}Shanghai Engineering Research Center of Digital Educational Equipment, East China Normal University, Shanghai, China  
\*xqgu@ses.ecnu.edu.cn

Abstract: Ongoing feedback from teacher to pupil during the instruction and interaction play an important role for pupil’s learning. In a mobile computer supported teaching environment, the assessment and feedback are often designed to be in a concise and simplified manner for ease of use and data recording. In this paper, we aim to understand the nature of the ongoing feedback in this setting. By employing principal component analysis on a three-semester period dataset, we found the ongoing feedback of the teachers to the pupils can be decomposed into assessment oriented component and encouragement oriented component. The teachers tended to use more assessment oriented feedback than encouragement oriented as they went through from semester 1 to semester 3.

Keywords: iPad; ongoing feedback; classroom behavior management system

1. Introduction

Ongoing feedback from teacher to pupil during the instruction and interaction, in the process of formative assessment, is a prime requirement for pupil’s learning progress. Through the timely feedback, teacher convey their judgment on pupil’s current state or performance that is intended lead to improved performance and also extend their care and attention to pupils. From these feedback, the pupils are motivated to accommodate themselves to adjust to the formative assessment system and to fit well in it, especially for young children (Brophy, 2013; Evans, 1996). The orientations of teacher’s assessment are presented in the ongoing feedback and thus powerfully influence the pupil’s motivation, learning achievements.

Despite its generally recognized power, there is considerable variability on effectiveness of feedback, which implies that some types of feedback are more effective than others (Kluger & DeNisi, 1996). Existing empirical analysis and theoretical framework mainly focus on traditional oral assessment and feedback from three aspects: 1) what should be primary key features of effective feedback, e.g. Hattie and Timperley (2007) and Tunstall and Gipps (1996); 2) what make feedback effective or ineffective in real educational setting, e.g. Nelson and Schunn (2009) and Lu and Law (2012); 3) how to give effective feedback to pupils, e.g. Brookhart (2008) and Johnston (2004). However, little attention has been given to teacher’s ongoing feedback in mobile computer supported teaching environment. To keep them easy to use for the teachers, these computer-based tools are often designed to be as concise as possible and the teachers are required to accommodate their teaching to the environment (CHEN & GU, 2016). Exploration the nature of feedback in these technology enhanced teaching environment would provide us a new start point for further improvement (Lee, 2009).

The main purpose of this paper is to explore the main factors the drive the teachers’ ongoing feedback in a primary classroom via an iPad-based classroom behavior management system (CBMS). The paper is organized as follows. We first review related literature and develop our hypothesis in following section 2 and 3. We then describe our data in section 4. Following that, we test our research hypothesis in section 5. Finally, the paper concludes with practical implication of our results.
2. Literature Review

Feedback is critical for learning any new skills because it provides the pupils or teachers about his or her performance information that is intended to lead to improved performance (Chan, Konrad, Gonzalez, Peters, & Ressa, 2014). Although giving feedback is a generally accepted practice in educational settings, the nature of the feedback and the context in which it is given matter a great deal to its effectiveness (Brookhart, 2008; Mory, 2004) and little consensus exist about what constitutes qualitatively good feedback (Kluger & DeNisi, 1996; Nicol & Macfarlane-Dick, 2006). To understand the nature of effective feedback, Hattie and Timperley (2007) reviewed the evidence related to the focus of feedback and categorized feedback into four operation levels. They concluded that feedback at the process and self-regulation level seemed to be most powerful to enhance pupil learning, followed by feedback at the task level. Feedback at the self-level, defined as personal feedback and unrelated to specific of the task, seems least effective to enhance learning, but is too often used in classroom situations (Hattie & Timperley, 2007). These four levels resemble the learning activities that were distinguished by Vermunt and Verloop (1999); namely cognitive, affective, and meta-cognitive learning activities (Van den Bergh, Ros, & Beijaard, 2013). Since the affective factor is always part of the interpersonal communication between teacher and pupil (Meyer & Turner, 2002), Bergh, Ros and Beijaard considered the affective feedback be embedded in the nature of feedback, instead of as a separate focus of feedback (Van den Bergh, et al., 2013). From a broader view, Nelson and Schunn (2009) differentiated feedback into cognitive and affective categories. Cognitive feedback targets the content of the work and involves summarizing, specifying and explaining aspects of the work. Affective feedback targets the quality of work and use affective language to bestow praise and criticism.

When the context of feedback was confined in primary school, Tunstall and Gipps (1996) developed a typology of teacher feedback based on observations. They also emphasized that we should see these types on a continuum rather than completely separate categories (Hargreaves, 2005). For example, the socialization oriented feedback underpinned other types of feedback and other types feedback contained their own social codes (Tunstall & Gipps, 1996). They also pointed out the social norms which underlie evaluative feedback were highly important and the range of issues coming into play were complex if we aimed to analyze specific feedback. Since feedback is a social activity, we can understand it only by taking account of the social, cultural, economic, and political contexts in which it operates (Gipps, 1999). Other researchers have concentrated on teasing out what makes feedback effective or ineffective (Lu & Law, 2012; Nelson & Schunn, 2009) and describing the characteristics of effective feedback (Johnston, 2004; Van den Bergh, et al., 2013).

3. Hypothesis and Method

Following previous discussion by Nelson and Schunn (2009), the feedback can be broadly differentiated into two categories: namely cognitive and affective. Based on this, we can first propose that the feedback of the teachers by iPad-based CBMS can be roughly divided into two components. The first component is assessment oriented which evaluate whether the pupil’s learning or behavior aspect in the explicit or implicit norms. The second is encouragement oriented which focus on using positive affirmation to encourage pupil’s effort. Besides Nelson and Schunn (2009), these types of orientation in traditional oral assessment and feedback are discussed in previous literature, e.g. Tunstall and Gipps (1996), Hattie and Timperley (2007) in different names. Based on these discussions, we propose our first hypothesis as following:

\[ H1: \text{Teachers’ rating of the pupils by iPad-based CBMS can be represented in two common factors.} \]

Since the assessment dimension often intends to reflect the actual pupil status and previous master level and habit of behavior tend to positively correlated subsequent learning outcome and behavior, we can propose a hypothesis that subsequent scores on the first principal component are positively correlated with the previous scores. Meanwhile, since the encourage dimension is often used to encourage pupils and its transformation into actual pupils’ good outcomes depends on more
uncontrollable conditions (Hattie & Timperley, 2007), the correlation between the scores on this dimension is not consistently significant. Therefore, we propose the second hypothesis as following:

\[ H_2: \text{Subsequent scores on the assessment dimension are positively correlated.} \]

Since the indicators employed in our iPad-based CBMS were often very concise, the exact meaning of a specific rating data may contain mixed cognitive and affective oriented aspects (Van den Bergh, et al., 2013; Vermunt & Verloop, 1999) or in an evaluative-descriptive continuum (Tunstall & Gipps, 1996). To distill these mixed components, principal component analysis can be employed (Johnson & Wichern, 2007).

4. Data

The data were collected from four primary school classes in East China range from September, 2013 to January 2015, across three semesters. In this period, four different Chinese teachers used an iPad-based CBMS to rate the pupils’ classroom behavior and learning performance on several predetermined concise indicators. The rating data served as an ongoing feedback to the pupils and was recorded in the backend data management module. Figure 1 depicts a screen of the CBMS app, in which each avatar image indicates each pupil. When a teacher interacts with a specific pupil, he or she can access and rate this pupil’s performance on the predetermined behavior indicators by clicking the pupil’s avatar. If the pupil performs well on the indicator(s) of the current context, then the teacher can rate her or him positively. To reduce the workload additionally introduced by the iPad-based CBMS for the teachers, the behavior indicators were selected as concise as possible.

![Figure 1. User interface of the CBMS app on iPad.](image)

| Table 1: Classroom behavior assessment indicators in three semesters. |
|------------------------|-------------------------------------------------|
| **Semester**           | **Indicators & Abbreviation**                   |
| Semester #1            | (1) Listen to teachers attentively (LTA1);     |
|                       | (2) Speak clearly (SC1);                       |
|                       | (3) Response to teacher’s question actively (RTA1); |
|                       | (4) Recognize and read new Chinese characters correctly (RCC1); |
|                       | (5) Copy Chinese characters correctly (CCC1);  |
|                       | (6) Participate the game actively (PGA1).      |
| Semester #2            | (1) Listen to teachers attentively (LTA2);     |
|                       | (2) Spelling Chinese characters correctly (SCC2); |
|                       | (3) Comprehensive expression (CE2);            |
|                       | (4) Read the sentences correctly (RSC2);       |
|                       | (5) Copy the Chinese characters standardly (CCS2). |
| Semester #3            | (1) Listen to teachers attentively (LTA3);     |
|                       | (2) Read the sentences correctly (RSC3);       |
|                       | (3) Compose complete sentences (CCS3);         |
|                       | (4) Writing Chinese characters correctly and properly (WCP3). |
Four Chinese teachers in four first grade classes (with a total of 124 pupils) participated in the study between September, 2013 and January 2015. The enrollment in these four classes was 34, 28, 34 and 29 respectively. The predetermined learning and behavior indicators for individual semesters are listed in Table 1. In total, 27,227 pupils’ performance data points were collected during the three-semester study period. Each data point represents a pupil’s performance on one particular indicator in one day.

5. Analysis and Results

Principal components analysis (PCA) was conducted on the rating data to test $H_1$ (Note that the common factors are equal to principal components if we employ PCA-based approach to infer factors, see Johnson and Wichern, 2007). Results are shown in table 2. It was found 8 out of the 9 percentage of total rating variances explained by the two principal components are over 80%. Note that the teachers were not familiar with the new tools in the first semester and the randomness in their rating behaviors was more than the other two subsequent semesters. The results indicate that two components are sufficient to represent the variances in the original indicator-rating data. Hypothesis $H_1$ is supported by results in table 2.

Table 2: Proportion of explained rating variance by the first two components in three semesters.

<table>
<thead>
<tr>
<th>Explained variance by the first two components</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Semester</td>
<td>69.57%</td>
<td>81.41%</td>
<td>87.47%</td>
<td>81.55%</td>
</tr>
<tr>
<td>Second Semester</td>
<td>91.77%</td>
<td>91.98%</td>
<td>97.13%</td>
<td>91.78%</td>
</tr>
<tr>
<td>Third Semester</td>
<td>95.28%</td>
<td>95.95%</td>
<td>96.20%</td>
<td>94.85%</td>
</tr>
</tbody>
</table>

To test $H_2$, the correlations between scores on the previous two principal components and subsequent components across three semesters are analyzed by linear regression. The R-square ($R^2$), estimated coefficient (coef), standard error (se) and p-value are reported. We used abbreviation “SX_COMY” to stand for “ Component Y in Semester X ”. The results are shown in figure 2. From figure 2, we can see that the assessment component in the first semester is positively correlated with the assessment component in the second semester and this positive correlation hold the same between the second and the third semester. Note that the $R^2$ between the first semester and the second semester are generally smaller than the $R^2$ between the second and the third. This means more assessment and feedback on the learning and behavior dimension are included in the feedback. This can also be verified from two other points. First, the teachers added more academic related indicators in their ongoing assessment and feedback, as in table 1. Second, for the four classes, the proportions of the component “assessment” in the two components generally increased from the first semester to the third semester, as shown in figure 3.

![Figure 2(a). Correlation of the two components in class 1.](image-url)
6. Discussion and Conclusion

By the study of a three semester feedback data collected by an iPad-based CBMS, we found we can interpret the ongoing feedback of the teachers from two independent viewpoints, namely assessment orientation and encouragement orientation. The assessment oriented component mainly evaluated the pupils’ learning or behavior aspect, while the encouragement oriented component focused on positive affirmation to encourage pupil’s effort. These two factors can be mixed in actual teacher ratings on specific behavior indicator. Based on this viewpoint, we also found the teachers tended to encourage the pupils’ by using their feedback from the beginning. When the teacher got along with the pupils for longer time, the teachers tended to use more assessment oriented feedback to evaluate pupils’ actual learning behavior or performance.
Our study shed light on how to explain the collected feedback data in a teacher rating manner using iPad-based behavior management system. Our finding has several implications for teachers and educational researchers interested in designing and implementing concise electronic feedback for pupils. First, teachers can use the extracted component from their historical feedback data to reflect their assessment on the pupil. It is often not accurate for the teachers to reflect their teaching activities solely based on their raw recorded feedback data. The decomposed data would provide the teacher a relatively objective base to reflect their teaching activities. Second, the two confirmed components help the educational researchers to explore the validity of the cognitive assessment component and the affective encouragement component. Since the original feedback data are often quite complex caused by teachers’ personal pedagogical experiences, principal components of the original feedback data provide us a more straight way to validate the effectiveness of the feedback by their actual meaning. In future study, we suggest to investigate how the two components affect pupils and under what conditions or circumstances.

Acknowledgements

This work was supported by the Key Project of Science & Technology Commission of Shanghai Municipality of China (Grant No.: 17DZ2281800).

References

Brookhart, S. M. (2008). How to give effective feedback to your students: ASCD.
Development of a Community-based Hazard Information Sharing System for Traditional Towns with Local Heritage

Shun KOZAKI*, Yasuhisa OKAZAKI*, Hiroshi WAKUYA*, Yukuo HAYASHIDA*, Byung-Won MINb & Nobuo MISHIMAa

Saga University, Japan
Mokwon University, Korea
*s-kozaki@ai.is.saga-u.ac.jp

Abstract: This paper describes development of a community-based hazard information sharing system by local residents in their daily life. We designed and developed an unique ICT-based hazard information system which contributes to community based disaster prevention and reduction. The continuous resident participation and posting design are core concept for our sustainable community-based approach. Our system continues to support making a hazard map by integrating the community-based hazard information. Local residents register information about the spot that can be dangerous in case of disaster. In addition, this system enables information sharing through a Web server. We expect that this information sharing effects usefulness of our system by utilizing collected local hazard information of each district.

Keywords: disaster prevention, hazard map, traditional town, ICT-based, community-based

1. Introduction

Japan has many natural disasters such as earthquakes and typhoons, or volcanic eruption. Therefore, studies on these disasters have been made widely. Not only a DIG (Disaster Imagination Game) which is a map exercise for disasters to enhance disaster prevention of communities(Komura & Hirano, 1997), but also a variety of ICT-based approach including information collection by using cameras and sensors, notification of disaster information by using ICT or game-based training has been proceeded (Mitsuhara et al., 2015). The traditional local towns are characterized by preservation of traditional landscape and environment, depopulation and aging. They are vulnerable to disasters because these factors causes spatial and human constraints.

Our approach in this paper is small start scalable ICT-based disaster prevention, which rooted in the region and based on the characteristics of these towns as disaster prevention efforts in these constraints (Okazaki et al., 2016). The traditional local towns are a characteristic that there is a strong connection between residents. Our idea is to utilize the power of these local communities in disaster prevention and reduction by ICT. To achieve safety and sustainability of livelihoods, our system encourages the local residents to be conscious of disaster risks and to participate disaster prevention/reduction activities. Furthermore, it enables the detailed hazard information collecting by local residents themselves. This continuous resident participation and posting design can make a significant contribution to sustainable community based disaster prevention and reduction.

We selected Hizen-Hamashuku in Kashima City of Saga Prefecture in Japan as a model areas of traditional local towns. We have tested the usefulness and possibilities of our prototype system in the model area (Okazaki et al., 2016). The easy-to-use interface contributed to the smooth registration of information. Our pilot study suggests usefulness of our ICT-based hazard mapping by local residents.

In this study, we designed and implemented a community-based hazard information sharing system for traditional towns. By this system, we expect that residents have higher knowledge of disasters and deeper awareness of disaster prevention.
2. System Overview

We have developed a hazard map creation support system with community participation type using the location information. Disaster prevention awareness can be improved by participation activities. Local residents can collect detailed information. Making hazard map with resident participation activities can improve sharing local community information. The exchanges of the conventional information based on conversations, telephones and letters. This system consists of three functions, information posting function, information sharing function and information management function. Using these functions, this system makes it possible to collect and provide information from residents.

3. System Functions

3.1 Information Posting Function

Information posting function is composed of a user selection screen, a map screen, a positional information screen and an information registration screen. The map screen displays risky locations stored in a database and the present location of the user acquired by GPS. The balloons point to the risky locations. The photograph of the location is included. A user can watch the information (a disaster type, a risk level, comments) of the location by tapping the balloon. On a positional information screen, users designate the location that users feel danger. When a user registers information, at first user drags a pin with a positional information registration screen and appoints a risky location. Then the position data that the pin points at is handed to the next information registration screen. On an information registration screen, this system saves information such as a disaster case, the photograph of the location and risk level in SQLite of tablet-type devices. Saved information is displayed on a map screen.

3.2 Information Sharing Function

Information sharing function is available to tap the information update button on the map screen in this system. Then user tap the button, saved information in SQLite of tablet-type device is sent a designate
PHP program on the Web server. This sent information is taken off the information sent before. This sent information is saved in MySQL in the Web server. The photo data is sent a directory in the Web server and the reference path of photo data is saved in MySQL. Consequently, information is collected from each tablet-type device and integrated in the Web server. The integrated information is output as a JSON file. Each tablet-type device gets that file. Received information is overwritten before saved information in SQLite of tablet-type device.

3.3 Information Management Function

Information management function has four functions, they are an approve function, a display function, a modify function, and an integrate function. The aim of the approve function is to ensure reliability of information that has been posted by system administrators to approve information. Display function is aimed to provide only the information necessary for the residents by hiding the old information and the information that is not necessary to be displayed among posted information. Modified function is aimed to provide more reliable information by putting function to modify the posted information on the Web server side. The integrate function is aimed to improve a quality and ease of viewing of the information by organizing and integrating a number of information. The integrated information is directed to the information that is several posted to the same location. The integrated information is newly registered in the Web database. As we can refer to the information prior to integration, the system keeps integration history.

4. Concluding Remarks

We have developed a community-based hazard information sharing system for traditional towns. Sustainable community-based disaster prevention/reduction is quite characteristic of our approach. We are focusing safety and sustainability of livelihoods. Our system encourages the participation of local residents and collects the detailed hazard information of the area by post of residents themselves. The continuous resident participation and posting design are core concept of our approach, which can make a significant contribution to sustainable community based disaster prevention and reduction. As future works, we will establish the management structure of the system. Moreover, we need to try this enhanced system in the field and evaluate it. After evaluating the system, we are going to apply our system continuously on a larger scale to demonstrate usability of our ICT-based modern approach of community-based disaster prevention and mitigation.

Acknowledgements

This study is supported by the funds of Japan Society for Promotion of Science (JSPS) and National Research Foundation of Korea (NRFK)’s bilateral Joint Research Projects during 2014 to 2016, and by JSPS KAKENHI Grant Number JP16H04478. We would like to thank persons resident in Hizen Hamashuku and all who understood and cooperated our on-site field work. We also wish to appreciate Mr. Hidaka's great help and valuable discussions with all project members.

References

Educational Effectiveness of a System for Scientific Observation of Animals in a Zoo

Yui TANAKA⁎, Ryohei EGUSA⁎⁎, Yuuki DOBASHI⁎, Fusako KUSUNOKI⁎, Etsuji YAMAGUCHI⁎, Shigenori INAGAKI⁎ & Tomoyuki NOGAMI⁎

⁎Kobe University, Japan
⁎⁎Research Fellow of Japan Society for the Promotion of Science, Japan
⁎⁎⁎Tama Art University, Japan
⁎⁎⁎⁎163d305d@stu.kobe-u.ac.jp

Abstract: Tanaka et al. (2017) developed a system using animation to help children observe the anatomy and behaviors of animals in zoos. A zoo is an informal learning place providing children a chance to observe animals scientifically and systematically. However, most zoos fail to support the observation of animals, and hence there is a need for more research on supporting observation in zoos. Animation is a useful tool to support observation of animal behaviors. The system developed by Tanaka et al. (2017) supported children in noticing not only the outward features and behaviors of animals but also their anatomies and movements. A workshop was held at Kobe Municipal Oji Zoo in Japan, in which children used the system to observe how penguins swim and walk and to note the features of the bones in their legs and flippers. The participants were 19 children, average age 8.1 years (SD = 0.6). In this paper, we examined the percentages of children’s scientific answers to questions incorporated in the prediction, observation, and subsequent interview one month after the workshop. There were four questions on swimming, walking, leg bones, and flipper bones. Each question had three choices. The percentage of scientific answers during prediction, observation, and interview was different for each item. This system was effective for some items, but not for all. In future work, we need to improve the contents and the animations of this system.

Keywords: Zoo, scientific observation, animation

1. Introduction

One of the places of informal science learning is the zoo, where children can observe animals in their natural habitat and become interested in nature (National Research Council, 2009; Falk, 2014). Eberbach and Crowley (2009) point out that observation requires the ability to organize phenomena into scientifically meaningful patterns, and as a preliminary step toward learning scientific observation, children should be able to connect features of animals to their function and behavior. Patrick and Tunnicliffe (2013) said that zoos should help children observe the real-life correlates of simple morphological and taxonomic terms. They also note that taxonomic judgments about organisms cannot be made without understanding their anatomy and behavior. However, children lack knowledge of such matters and need support to notice those (Prokop et al., 2007). In addition, zoos are generally considered a place for leisure; so, most visitors come only to look at the animals and be entertained by them (Clayton et al., 2009). Some of the educational programs in zoos may not impart much knowledge about animals’ habits, and therefore research is needed to develop programs that support learning about animals (Patrick & Tunnicliffe, 2013) to make zoos an effective place for science learning.

Tanaka et al. (2017) developed a system using animation to help children observe the distinctive anatomical features and behaviors of penguins. Technology-based education in a zoo has a greater impact than static signage and attracts the attention of young people more easily (Webber et al., 2016). Animation helps children visualize and motivates them to learn the movement of an animal (Betrancourt & Chassot, 2008; Patrick & Tunnicliffe, 2013; Shreesha & Tyagi, 2016; Tanaka et al., 2016), its outward features, as well as its internal structure.
The aim of this paper, therefore, is to examine the effectiveness of this system, by making children observe and learn about penguins and their skeletons. Finally, a comparison is made of the percentage of scientific answers to questions asked at prediction, observation, and interview phases.

2. Methodology

2.1 System Overview and Workshop

Tanaka et al. (2017) made up four multiple-choice questions on the anatomy and behavior of penguins, respectively covering swimming, walking, leg bones, and flipper bones (Figure 1). Each question has three answer options, which were emphasized for children to gain a better understanding.

To ascertain the effectiveness of the system, Tanaka et al. (2017) held a workshop using the system at Kobe Municipal Oji Zoo, in Japan, on December 1, 2016. The participants were 19 children, average age 8.1 years ($SD = 0.6$). One tablet running the system was provided to each child.

First, children were asked a question about penguins’ anatomy or behavior. Children watched the animations for each choice and chose one prediction. Following this, they observed the actual penguins or the skeleton of a penguin; they also watched an animation during the observation phase. They then selected the answers from among the choices once again based on the observation. Last, the staff members explained the feature or behavior with scientific details.

![Figure 1](image-url). The page about flipper bones of the system, by Tanaka et al. (2017).

2.2 Evaluation

For evaluation, the percentage of scientific answers children provided during prediction and observation was analyzed. A month after the workshop, the children were interviewed. They selected a scientific choice relating to the anatomy or behavior of penguins and explained it.

3. Results

Table 1 shows the percentage of scientific answers to each question. On swimming, the percentage of scientific answers in the observation phase (100%) was higher than that during prediction phase (63%); the interview rate was 47%. On walking, the prediction rate was 95%, observation was 100%, and interview was 95%. On the leg bones, observation (95%) and interview (82%) results were higher than the prediction phase results (16%). On flippers, observation phase results (32%) were lower than prediction phase results (53%), but the interview results (78%) were higher.

<table>
<thead>
<tr>
<th>Items of questions</th>
<th>Prediction</th>
<th>Observation</th>
<th>Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swimming</td>
<td>63%</td>
<td>100%</td>
<td>47%</td>
</tr>
<tr>
<td>Walking</td>
<td>95%</td>
<td>100%</td>
<td>95%</td>
</tr>
<tr>
<td>Bones of legs</td>
<td>16%</td>
<td>95%</td>
<td>82%</td>
</tr>
<tr>
<td>Bones of flippers</td>
<td>53%</td>
<td>32%</td>
<td>78%</td>
</tr>
</tbody>
</table>
4. Discussion and Conclusion

This paper described the evaluation of a system using animation to support children’s observations at a zoo, specifically, to observe penguins’ anatomy and behaviors. The effectiveness of this system was different for each item observed. On swimming, children related to swimming using legs and about half of them gave a non-scientific answer at prediction and interview phases. By using this system, children could temporarily change their thinking. It implies they have a misconception that penguins swim using legs. In order to change this misconception, there is a need to improve the system, for example, by asking them to compare the motions during swimming by the penguin to the flight of birds. On walking, the prediction rates and interview rates were high implying this knowledge was not because of the system. Regarding the bones of the legs, the rates at observation and interview were higher than at prediction, proving the effectiveness of the system. Regarding the bones of flipper, the rate at observation was lower than that at prediction. This is because the options were emphasized and therefore, children found a difference between what was emphasized and what was real. It created confusion in the minds of the children. However, one month later, children gained knowledge about bones on the flippers, and hence the system was considerably effective. In future work, we need to improve the contents and animations of this system for each item.

Acknowledgments

This research was supported by JSPS KAKENHI Grant Number 16H01814. We are grateful to Kobe Municipal Oji Zoo, which made this study possible.

References


Leveraging an Existing Learning Management System for Alternative Learning

Mikaela MALIT* & Ma. Mercedes RODRIGO
Department of Information Systems and Computer Science, Ateneo de Manila University, Philippines
*mikaela.malit@obf.ateneo.edu

Abstract: Alternative learning is a practical option for interested learners that have not attended or finished formal basic education. In the Philippines, the Department of Education offers many alternative learning and non-formal programs for every kind of Filipino. While there have been initiatives to integrate information and communication technologies (ICT) in these programs, the use of modern ICT has not been fully realized. This study investigated the effectiveness of introducing an existing learning management system (LMS) as a supplement to the Accreditation and Equivalency Alternative Learning System (ALS) program of the Department of Education. The LMS was intended to aid learning and improve performance by enabling online access to learning materials, delivery of real-time feedback through self-checking interactive quizzes, and self-monitoring of learning progress. A dynamic quiz generation plugin was developed and integrated with Moodle to create interactive quizzes from existing static learning materials. LMS use was monitored in an eight-week implementation in two learning centers where learning gains, usability, access, response, and overall experience were assessed. Based on the analysis of gain scores between the experimental and control groups, the users of the LMS performed better in the posttest. Evaluation of the LMS yielded favorable ratings in usability, and access and response as an online learning environment. Overall, the student and teacher respondents viewed the LMS as a helpful learning tool and encouraged its use in other learning centers and in the future. Feedback and suggestions were also gathered from the respondents during evaluation which can be used to improve the LMS.

Keywords: Applied computing, learning management systems, human factors, Department of Education of the Philippines, Alternative Learning System, Moodle, system deployment

1. Introduction

The Accreditation and Equivalency ALS program of the Department of Education of the Philippines is a non-formal education program that provides opportunities for out-of-school youth and adults to complete their basic education. Steps have been taken to introduce ICT to ALS and one is the Radio-Based Instruction program, a variety of distance learning that uses radio broadcasts to air learning modules in very remote areas. (Acido, Muega, and Oyzon, 2013) Another ICT initiative was Project MIND which explored how using SMS can reduce dropout rates and improve performance (Ramos, Librero, Triñona, and Ranga, 2007). A blended-type e-learning program was also launched in 2007 called the eSkwela project to promote ICT use. (UNESCO Bangkok, 2009).

The ALS programs have started using ICT but most if not all of these initiatives have ceased operation. Most ALS programs also use non-interactive learning materials such as booklets and handouts in teaching. There is an opportunity to capitalize on the interactive capabilities of modern ICTs and reintroduce e-learning. The goal of this study is to adapt the current static learning materials of the ALS programs and convert them into interactive learning tools that can be integrated into a modern LMS. The LMS is expected to improve student performance by enabling wider and more convenient access to online self-assessment learning materials at the students’ own pace, location, and time. The LMS will provide structure, content, and monitoring and assessment capabilities in the ALS programs that will help establish a rich learning experience for the students.
2. Research Questions

This study intends to answer the question “How should an existing learning management system be used to be an effective supplement to ALS sessions?” through the following questions: (1) What is an appropriate online learning environment for the ALS materials? (2) What functionalities of the chosen learning management system will be used and developed? (3) How will the student progress be monitored? (4) What measures will be used to determine the effectiveness of the learning management system? Based on these measures, how effective is the learning management system? (5) How will the learning management system be evaluated? (6) What is the overall students’ experience in using the learning management system? (7) How will the learning management system be sustained after its deployment?

3. Results

What is an appropriate online learning environment for the ALS materials? Moodle is an LMS that uses shared libraries, abstraction, and Cascading Style Sheets, and was developed with a focus on the elements of the students' learning environment (Dougiamas and Taylor, 2003). Moodle was chosen in this study because it is: (1) Free - publicly available with no licensing fees; (2) Open source - highly portable, extensible, and adaptable; (3) Scalable - easily upgradable therefore regular maintenance is not needed; (4) Stable - has a 14 year history, assuring up-to-date documentation and support; (5) Intuitive - largely intuitive and can be mastered quickly; and has a (6) Community - that develops and improves features.

What functionalities of the chosen learning management system will be used and developed? The standard distribution of Moodle was used with an additional checklist plugin. The Moodle Mobile application was also utilized. A dynamic quiz generator was developed based on open source programs: pdf2htmlEX and SCORM API wrapper. Pdf2htmlEX is a PDF to HTML converter and publishing tool that allows the content to be displayed and uploaded in the web browser as originally formatted and styled (Wang and Liu, 2013). Using pdf2htmlEX, ALS PDF files were converted into identical HTML files. The files were then adapted into interactive quiz sections through the SCORM API wrapper, an open source API that enables conversion of HTML files to SCORM packages that can connect to any SCORM-compatible LMS (Hutchison, 2008). The converted quiz SCORM packages enabled saving of the quiz scores in the Moodle gradebook.

How will the student progress be monitored? Student progress was recorded for eight weeks through four one-hour sessions in the learning centers and the LMS logs. A total of 50 students and 2 teachers from two ALS learning centers participated in the study from varying age groups (15-66 years old). Student participants were randomly divided into experimental and control groups.

What measures will be used to determine the effectiveness of the learning management system? Based on these measures, how effective is the learning management system? LMS effectiveness was determined by comparing the learning gains of the experimental and control groups in written tests before and after LMS use. For the experimental group, frequency of use and logins, number of quiz attempts, and quiz and module completion rate were also used. Based on learning gains, the experimental group (M=4.2 SD=17.75 N=25) performed better than the control group (M=2 SD=13.92 N=25), t(47)=1.95, one-tailed p<=0.03. High performing students (M=25237 SD=117574523 N=13) were shown to have used the LMS more than low performing students (M=15520 SD=76135636 N=12) t(23)=2.48, two-tailed p<=0.02. No one was able to complete all uploaded modules but it was found that high performing students (M=2.42 SD=1.16 N=13) completed more quizzes than the low performing students (M=0.77 SD=0.17 N=12) t(16)=5.11, two-tailed p<=0.0001. Based on these data, there is significant evidence that the LMS use in ALS can aid learning and performance.

How will the learning management system be evaluated? Two evaluation tools were used: the System Usability Scale (Brooke, 1986) and the Web based Learning Environment Instrument.
Students rated the LMS a 75.5 in the SUS which is a Good rating according to SUS grade rankings. However, this score is only a C which indicates that LMS usability can be improved. On the modified WEBLEI questionnaire, students had a self-report measure of 3.95 on Access. This indicates that the LMS provided above average access of learning materials. On Response, students measured 4.35 which shows that the LMS promoted and enhanced learning. These evaluation scores are indicative of how effective the LMS is to the ALS programs.

What is the overall users' experience in using the learning management system? Overall LMS experience was gathered through a self-report and opinion questionnaire. Students reported to have a positive experience based on their answers in the questionnaires. Users also provided suggestions that can be used for LMS improvement. Overall, the students and teachers positively viewed the introduction of the LMS and were generally hopeful of future LMS versions.

How will the learning management system be sustained after its deployment? Technical, personnel, administrative, and operational aspects of the LMS were discussed. Moving forward, LMS operation will continue with support from the partner learning centers in this project.

4. Conclusion

In this study, the effectiveness of introducing Moodle to the ALS programs in the Philippines was explored in an eight-week implementation in two learning centers. In general, the LMS was positively received by the students and teachers based on evaluation and feedback. There were also substantial learning gains from the students’ LMS use. As part of future work, the use of Filipino materials in the LMS and their effect on student performance are worth exploring. LMS use in this study was only implemented in two centers; therefore, deployment in a much wider scale will further determine the actual effectiveness of the LMS for alternative learning in the Philippines.

Acknowledgements

We extend our gratitude to the students and teachers of the partner ALS learning centers and the Ateneo de Manila University for their support of this study.

References

Providing Regular Assessments and Earlier Feedback on Moodle in an Introductory Computer Science Course: A User Study

Natalia NEHRING, Simon DACEY & Nilufar BAGHAEI
Department of Computer Science, Unitec Institute of Technology, Auckland 1025, New Zealand
nnehring@unitec.ac.nz

Abstract: Formal feedback on assessments is given to students half way through the semester for a lot of tertiary level courses. For some students, it can be too late to change their study habits and they might not have realised that they were struggling before getting the feedback. Receiving late feedback can also result in lowering learner’s motivation as well. The aim of this study is to determine the effect of providing regular feedback on students’ academic achievement. The sample group included 92 students from semester 2 2016 and 57 students from semester 1 2017. They were the first-year Bachelor Computer Science students at UNITEC Institute of Technology, New Zealand, doing an introductory database course. The course had three assessments in 2016 including two computers based tests and a final exam on paper. Weekly quizzes were introduced in 2017 with the aim of providing early regular feedback to students. Preliminary results show that providing this feedback early on did not significantly improve the results of their first formal assessment. More studies are needed to examine the effectiveness of weekly quizzes on students’ performance throughout the semester. Subjective evaluation showed that majority of students liked getting regular feedback in form of quizzes and found it valuable for their learning.

Keywords: regular feedback, designing quiz, quiz as formal assessment, LMS

Introduction

Usually first formal assessment takes place after six weeks of delivering the course materials in a lot of tertiary institutions, and then another week or two before students get any feedback on the assessment. For some students, it can be too late to change their study habits and they might not have realised that they were struggling before getting the feedback. Receiving late feedback can also result in lowering learner’s motivation as well (Auvinen et al, 2015). Studies have recommended routine quizzes to be introduced in any course in which objective learning goals need to be assessed. This is to improve students’ motivation and their self-assessment capabilities. (Balter, et al, 2013; Kenis, 2011). For example, Balter, Enstrom, and Klingenberg (2013) reported that that the students’ results had improved after introducing early feedback in form of quizzes.

Online learning management systems (LMSs) are extensively used by tertiary institutions for course delivery and are highly popular (Conde et al, 2014). One of the widely used open source platforms is Moodle (www.moodle.com). It is an open source platform that provides a learning space for users to design online courses with flexible content and collaborative activities. Studies have shown that using LMSs including Moodle can enhance students’ performance and/or satisfaction (Smith, 2016; Paechter et al, 2010) due to self-regulated and collaborative learning tools they are provided with.
The course we picked for this study is ‘Introduction to Database’ – it is offered every semester to first year Unitec students enrolled in a Computer Science degree. It has three assessments including two computer-based tests worth 25% each, and a theory paper-based exam worth 50% of the final grade. With 2016 cohort, the first test happened during week 6, with overall length of the course being 13 weeks.

For 2017 intake, we made some changes to the assessments in that 10% of the total grade was assigned to 10 newly developed quizzes, thus making each quiz worth 1%. Our experience shows that allocating even 1% of total grade to an exercise can encourage students to take it more seriously without worrying them about the assessment weight. Our hypothesis was that providing earlier feedback on Moodle could improve students’ learning outcome and encourage them to stay on track.

Research Questions

The research questions we are investigating in this paper are: 1) will regular assessments/quizzes (each worth one percentage of total course grade) improve the learning outcome of our students? 2) What is student perception of doing Moodle quizzes on a weekly basis?

Method

Moodle provides the ability to set up assessments using quizzes. For this project, we developed 10 multi-choice quizzes. Each quiz was set to have one attempt only and had 5 to 10 questions. Quiz questions were based on the previous week’s homework and tutorials, which students had been asked to complete in the own time. Quiz time was set from 15 to 20 minutes depending on the number of questions. The scores were shown to students after quiz completion.

A subjective evaluation was also conducted to find out what students think about quizzes. Questions about the user experience were developed, three of which were taken from Balter (2016) research.

Results

For 2016 intake, Test 1 was taken by 92 students. As semester 1 2017 is still in progress, only preliminary results can be reported. The same test was completed by 57 students. The average mark for the test one this semester was 67.4 (with sd 29, CV 43), which is slightly more than the previous semester 66.9 (with sd 25, CV 38.3), but the difference is not statistically significant.

Our hypothesis was that there is a correlation between quizzes marks and the main assessment’s marks and it would help to predict results. The results for test one compared to the average quizzes marks show that the standard deviation is 12. For 35% of students, marks are predicted (considering that the difference between test one and the average marks of quizzes is less than 10%). If we consider a 20 mark absolute difference or less, prediction can be done for 61% of students.

The subjective evaluation survey was done in week 9 after students completed 8 quizzes. The number of participants was 43, 74 % from all students enrolled in the course. Most students strongly believed that doing quizzes on a weekly basis improved their study habit (see Figure 1) The average answer is 8 out of 10.
Figure 1. Response to “Do you believe that every week quizzes affect your study habit?”

Table 1: effect of doing quizzes on study habits. Number and percentage of responses for each option available in Question 1, compared with previous studies described by Balter (2013)

<table>
<thead>
<tr>
<th>Option</th>
<th>KTH 2009</th>
<th>KTH 2010</th>
<th>Williams 2010</th>
<th>Unitec 2017 (existence of quizzes)</th>
<th>Unitec 2017 (quizzes results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, I studied harder</td>
<td>25 (25%)</td>
<td>11 (9%)</td>
<td>3 (10%)</td>
<td>19 (44%)</td>
<td>24 (56%)</td>
</tr>
<tr>
<td>Yes, I studied less</td>
<td>2 (2%)</td>
<td>5 (4%)</td>
<td>1 (3%)</td>
<td>4 (9%)</td>
<td>6 (14%)</td>
</tr>
<tr>
<td>No, I would have studied just as much/little without the tests</td>
<td>61 (60%)</td>
<td>67 (57%)</td>
<td>21 (72%)</td>
<td>14 (33%)</td>
<td>9 (21%)</td>
</tr>
<tr>
<td>I did not study more, but earlier</td>
<td>n/a</td>
<td>11 (9%)</td>
<td>2 (7%)</td>
<td>2 (5%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>I don’t know</td>
<td>14 (14%)</td>
<td>23 (20%)</td>
<td>2 (7%)</td>
<td>4 (9%)</td>
<td>3 (7%)</td>
</tr>
</tbody>
</table>

Conclusions & Future Work

In this paper, we investigated the effect of providing regular feedback on students’ academic achievement in an introductory database course. Subjective evaluation showed that majority of students liked getting regular feedback in form of quizzes and found it valuable for their learning. More studies are needed to examine the effectiveness of weekly quizzes on students’ performance throughout the entire semester.

References


Study on Implementing Automated Classroom Performance System for Recording Student Attendance

Daniel MANGALARAJ & Shankar SUBRAMANIAN
UNITEC, Auckland, New Zealand
*dmangalaraj@unitec.ac.nz

Abstract: We explore and study the implementation of an automated classroom performance system for recording student attendance. Classroom attendance is closely linked to student success. The traditional method of recording student attendance by signing on paper or by calling students names is time consuming and since it has the human intervention element it is insecure and hence inefficient. This paper introduces the replacement of manual registry system with automated computer based smart tag and reader. These customized tags and receivers are used to register a student’s presence in class. This information is later passed on to the administrator’s for admin follow up. The collected and stored data can be analyzed along with other student data using predictive analytics methodologies to predict student’s success in the program and also intervene at an earlier stage to help towards the success of the student.

Keywords: Automated attendance, student journey, Predictive analytics, monitoring attendance

1. Introduction

There is a strong connection between student attendance and their achievement. Absenteeism affects student performance and its ripples on to their educational journey. Sudheer et al. (2010) agree that from a teacher’s perspective, recording and collating student’s attendance for weeks to gather and analyse is a herculean task. In this project we intend to streamline the attendance register process and also change it from a manual process to automated computerised process.

The concept of using the IoT technology involves the use of sensors that can be tracked by base stations or readers positioned at strategic locations. Students and teachers will be issued with name badges that are smart and embedded with sensors that can be recognized by readers. The readers help to identify and track the presence of individuals at specific locations. This concept can be used for recording attendance of students in classes. Kassem et al. () have conducted a similar project, however their RFID tag does not monitor the attendance of students for the entire period in class, which means students could log in and walk out of the lecture.

Technologies available today for this application can be classified into two categories. (1) Scanning technologies and (2) Unobtrusive, Behind the Scenes technologies. Technologies that involve scanning the individual’s ID or badge at a scanning station form part of the former group while technologies that involve unobtrusive sensing of the individual’s presence in a location form part of the later. Examples of Scanning technologies include Bio-Metric Sensors, Bar Code Readers and Finger Print Readers. Examples of Unobtrusive, Behind the Scenes technologies work on sensors that are Blue Tooth, Wi-Fi, GSM or GPS enabled. These involve sensors that can be recognised automatically by readers. Also using a camera to scan individuals and recognising them using image recognition is part of this category of technology. RFID can belong to both categories depending upon how it is implemented. An RFID reader installed at the gate will automatically recognise an individual carrying an RFID tag. Alternatively, the individual can scan or touch his RFID tag on to a scanning device to register presence. Unobtrusive, Behind the Scenes technology has the advantage of detecting the continuous presence of an individual at a location. This will be particularly suited for recording student attendance. It requires minimum infrastructure changes as it can operate wireless.
2. Automated Attendance Model

In this section, we describe how this concept can be put to work for attendance recording. Each course has its own login. The faculty will be able to see all the students registered for the course. As students enter the class and occupy seats, their presence is sensed by the reader installed in the class room and the student’s attendance gets automatically recorded. Students present in the class will have a green blob besides their names and all other students will be in red. The system will also record the entry and exit times. Multiple entry and exit times can also be recorded. The attendance data can be automatically sent to a central administrator for further processing, reporting and storing. The attendance data collected can be used for multiple purposes including tracking student performance, identifying weak students with a view to providing them with more resources and guidance as well as guiding staff to tweak / change their teaching methods to make students more interested in attending class.

3. Automated Attendance Benefits

There are several benefits of using this technology for attendance recording:

- **Reduction in errors** – Risk of human error in recording attendance is reduced considerably.
- **Increase in productivity** - As the process is seamless and automatic, productivity increases and makes day-to-day operations more efficient and convenient.
- **Saves money** – Implementing a technology based automatic attendance solution with smart badges and readers will help to reduce labour costs.
- **Increase student and staff satisfaction** - Removes the drudgery of manually marking attendance for the staff and the students anxiety of ensuring their attendance is recorded.
- **Increase in security** – the smart badges can be used not only for tracking attendance but also for controlling access to restricted areas such as laboratories and other facilities including equipment such as photocopiers, scanners and printers.
4. Limitations of Automated Attendance Model

The following are two main limitations of this attendance recording model.

- **Technical limitations:**
  - Battery life – It is important for the battery to last for at least a year. Clever hardware and software design can enable to extend the battery life.
  - Sensing distance – this will usually not be a major limitation for most class rooms as it can be managed using a larger number of sensors or readers cleverly placed inside the class rooms.
  - Sensing of students outside the class room – there is always a possibility of the sensors picking up students who are just outside the class room. However such students can be identified and excluded as (1) they would not be among those enrolled in the course and (2) the detection of their presence would not be continuous and last throughout the duration of the class. Another way of dealing with this issue is through a clever design of the antenna on the sensors / readers which will only sense the tags present in a confined area.

- **Non-technical limitations:**
  - Students may not have the tags on them all the time. Such situations may be few and far between. In all such instances, a manual back up would be required.
  - Students may feel that they are being tracked. This is a false belief as the system is not tracking the student. It only indicates whether a student is located in a specific area or not at a given time; the area here being the class room.

5. Other systems for Automated Attendance Model

The other systems available for an automated attendance application are Bio-metric scanning, Facial recognition system using camera scans, Finger print reading, Bar code scanning and RFID scanning. All of the above except facial recognition require students to queue up at a scanner and have their badge or some part of their body (eye, finger) scanned. This will be time consuming and if it is a large class room, time spent in the queue would be wasteful. All of the above except facial recognition using a camera will not be able to keep track of the continuous presence of students in the class. They will only be able to track the entry of students into the class. As far as cost is concerned, bar code scanning may be cheaper than the rest.

6. Conclusions and Future Work

Once implemented this system would a) automate the entire process from recording students attendance until following up on non attendance, b) reducing fake attendance marking c)minimize teaching staff’s time on admin matters. Future developments could include trialing biometric finger print registry system, real time email notification to students on their absence and their subsequent monthly-customized email to students on their progress. Further study using predictive analytics to analyse student’s success along with other relevant student data would enable academics to intervene earlier with a view to help students succeed.

References


Abstract: As the complexity of distributed Internet of Things (IoT) applications increases, so does the need for better processing of raw data flow of sensors. And since many IoT applications require almost real-time reactivity to the environmental stimulus, such information inference process must be performed continuously and on-line. This paper proposes a semantic model of data flow processing and real-time reasoning based on Semantic Stream concepts as a natural extension of the Complex Events Processing (CEP) and RDF (graph-based knowledge model).

In this paper, we describe a scenario about patients flux monitoring in a clinic dental school. Finally, we will present perspectives and prospects on the new way of storage the Semantics Stream and finish the development of the Semantic Event Rules using ESPER EPL (Event Processing Language).

Keywords: Internet of Things (IoT); sensors; data streams; complex event processing (CEP); semantic reasoning

1. Introduction

With the expansion of the use of IoT applications using CEP (Complex Event Processing), the need to extend the processing of raw data arriving through sensors has arisen in the same proportion. These applications are demanded in various domains, including education. To generate an information or generate an action, it is necessary that this data can be treated in a more complex way than the common processing. Therefore, in order to improve this data processing, this process has been enriched semantically. The inclusion and storage of triple Resource Description Framework (RDF), provides a semantic enrichment of complex event processing (Adi, Botzera and Etzion, 2000). This semantic model enables the discovery of facts that were not possible to be extracted with simple processing (Teymourian, Rohde and Paschke, 2012). This article proposes the application of semantically enriched CEP rules to generate simple scalable actions.

With the goal of finding a suitable semantic model for IoT, this paper proposes a approach for real-time symbolic reasoning based on the concepts of Semantic Stream and Fact Stream, as natural extensions of Complex Event Processing (CEP) (Luckham, 2001) and RDF (graph-based knowledge model) (Candan, Liu and Suvarna, 2001). This article presents a general idea of IoT in the second chapter, a scenario of health education in the third, and finally, in the fourth chapter, the conclusion and future work of the proposal.

2. General Idea

Our proposal is to define a CEP transformation, through a semantic flow of events, in order to create information of a higher level, that is, semantically enriched information. The proposal foresees generation of RDF triples, from pre-processed events. Thus, it is possible to merge these triples and generate what we call the flow of facts. With this, we can detect the type of entity and if an action was generated by it. The identification of entities and predicate identification is performed by the CEP agent.
next to the sensors, (see Figure 1), which in the specific case of our IoT middleware usually runs on mobile devices. Therefore, we call them Mobile Event Processing Agents (EPAs). In Figure 1, it can be seen that through these agents, mentioned above, two of the three RDF triple data can be generated.

![Figure 1. Semantic annotation from raw sensor data.](image)

3. Scenario

Considering a clinical school dentistry environment in which several patients are targeted to a group of students. A common problem is that patients do not have adequate and continuous treatment, in addition to waiting to know if it is being referred. A general organization of patient flow is presented in Figure 2.

![Figure 2. Organization of the Dentistry school clinic.](image)

The reception is the first place where the patient will be register. Then he is sent to the screening sector, this sector is composed of an advisor along with the students of the second year of the course of dentistry. Later, upon detection of the problem, the student informs the reception of the current patient case and where it should be routed. Therefore, the responsibility to arrange the consultation with students who are in later periods, is with the reception. Depending on the diagnosis, the patient can be referred to the following sectors: i) Emergency (3th year dentistry students): Acute pain and dental extrusion; ii) Low complexity (4th year dentistry students): Restorations and scraping; and iii) Medium/High complexity (5th year dentistry students): prosthesis and endodontics.

3.1 Assumptions

To facilitate the understanding of our example, let us divide the explain up into two parts. Our goal here is to identify bottlenecks with the following characteristics: i) The patient spends a lot of time in the waiting queue; ii) Some students do not appear on the scheduled appointment day and iii) The patient is left unattended because he can not be relocated to another student who is available.

To infer these complex events, we can make some hypotheses: i) The patient and the student will have a mobile device in which an ID will be associated with identification; ii) The student will be allocated according to time in the course, that is, second, third, fourth or fifth year; iii) Each room will have a BLE sensor to identify how many and which students are present in the clinic; iv) Every patient chair will have a BLE sensor to identify whether it is occupied or not; v) The screening sector will have a BLE sensor to identify patients' arrival and departure flow.

3.2 Proposal
From the above premises, we can elaborate a CEP strategy for solving our problem. The sensors present at the reception will detect the arrival flow of the patients who are already previously associated with the students in charge of their cases. In addition to this, the system will be able to detect if the student in question is present. If the student is not present, the system will be able to relocate the patient to a new student at run time. Our proposal becomes clearer in Figure 3, where there is a flow of triple RDFs with CEP rules that are processed and stored in a knowledge base. This base is responsible for storing RDF triples in order to provide facts through the processing of Semantics streams.

![Figure 3. CEP nodes and information flow.](image)

4. Conclusion and Future Works

This paper presented a real-time reasoning approach based on semantic events for IoT applications. The proposal takes into account a scenario in which all objects, people, places, environments, etc. there are sensors in order to emit simple events whenever some action is performed to or with it by an actor and that each event will take UUID exclusive of the items and the precise time that has occurred. Restricting the predicates in a triple RDF in action how to get to, associate, attend, etc., rather than state, such as has, is, belongs to, etc., is limited the amount of information that flows Data / Events can express. However, it is believed that if we base all predicates on action we will raise the importance of reasoning in IoT applications. The next steps, we will finish the development of the Semantic Event Rules using ESPER EPL (Event Processing Language) and deploy them on our mobile IoT middleware.

Acknowledgements

Acknowledgements Our ESMOCYP (Efficient Semantic MOdels and Fault- tolerant Middleware for CYber-Physical Systems) Project is supported by PROBRAL CAPES–DAAD Brazil-Germany cooperation program (Process No 8148/2015-05) and by a CAPES PVE fellowship to J.-P. Briot.

References

A Meta Analysis: The Effectiveness of E-Schoolbag Use on Students’ Academic Achievement in China

Menghua Hu & Yiling Hu *
Department of Educational Information Technology, East China Normal University, Shanghai, China
* ylu@deit.ecnu.edu.cn

Abstract: This meta-analysis examines the effects of E-Schoolbag use on students’ academic achievement in China. From the previous studies data, the researcher calculated the overall, average effect size on students’ academic achievement, and analysis the effect in different subject and school level. The result of this study is as follows: the overall, average effect size of E-Schoolbag use to be 0.433, which is a small-to-medium effect size; The effect of E-Schoolbag varies from subject and school level. These findings indicate that E-Schoolbag have a positive effect on student learning to some extent. To improve the effect of E-Schoolbag, government, school and teachers should continue promote the integration of E-Schoolbag in the classroom.

Keywords: E-Schoolbag, Academic Achievement, Meta-Analysis

1. Introduction

With the rapid development of information technology and the popular of one to one digital learning, many digital learning tools have been developed to support students' mobile learning and ubiquitous learning. Many countries also have released policies, launched programs and improved investment to encourage and support schools to create mobile learning and ubiquitous learning environments, like the one to one iPad Program in American, the Digital Textbook's Commercialization Plan in South Korean. In China, schools use the Electronic Schoolbag (E-Schoolbag) to create the personal learning environment in the classroom for students.

E-Schoolbag is an important digital learning tool in the process of advancing and developing educational informatization in China. It is not a simple electronic material of traditional textbooks, but a personal learning environment that integrates electronic textbook learning resources, electronic textbook readers, virtual learning tools, digital learning terminals and connected ubiquitous learning services (Zhu & Yu, 2011). Now, E-Schoolbag has been widely used in Chinese primary and secondary school students’ classroom learning. However, the effect of using E-Schoolbag for learning is still difficult to define and the effect of using E-Schoolbag for learning is still have a debate. The purposes of this study are to examine how much influence the E-Schoolbag has on students’ academic achievement. At the same time, this study also considers the degree of E-Schoolbag’ impact on the student's academic achievement when E-Schoolbag are applied to different subjects and different school levels. According to research result, we can find out the impact level of E-Schoolbag on student learning and draw some suggestions for the E-Schoolbag's effective application in Chinese schools.

2. Research Methods

This study uses the method of meta-analysis to analysis 20 primary pilot studies of the effectiveness research of E-Schoolbag Use on Students’ learning. All those previous studies contain 33 effect size. The literature sample mainly search from those database, such as CNKI, Web of Science, Springer. The research topic of those samples should be the effect of E-Schoolbag in learning. And all those literature samples should consist of complete study data. This study used the Comprehensive Meta-Analysis (CMA) software to calculate all the effect size. After entering the data above, the CMA software automatically calculates the overall average effect size of all the studies. The standardized mean
difference (Cohen’s d) is used to describe the size of the effect, Cohen (1988) developed it for statistical analysis. The value of d is calculated as follows:

\[ d = \frac{X_e - X_c}{sd} \]

\[ sd = \sqrt{\frac{(n_e - 1)s_e^2 + (n_c - 1)s_c^2}{n_e + n_c - 2}} \]

X_e means the experimental group, X_c means the control group, sd is standard deviation. n_e is the sample size of the experimental group, and n_c is the sample size of the control group. s_e is the standard deviation of the experimental group, s_c is the standard deviation of the control group.

### 3. Results

The E-Schoolbag overall analysis result of the 20 articles selected for this study can be seen in Table 1. The overall, average effect size of E-Schoolbag is d=0.433. According to Cohen’s (1988) interpretation of effect size, when d is less than or equal to 0.2, it is considered to have a small influence. When d is greater than or equal to 0.5, it is considered to have moderate influence. When d is greater than or equal to 0.8, it is considered to have very significant influence. The result means that the E-Schoolbag have a positive effect on student learning to some extent, but is slightly lower than the medium effect.

Table 1. E-Schoolbag’s Overall Average Effect Size on Academic Achievement

<table>
<thead>
<tr>
<th>Model</th>
<th>n</th>
<th>Effect size and 95% confidence interval</th>
<th>Test of null (2-Tail)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>d</td>
<td>Lower limit</td>
</tr>
<tr>
<td>Random</td>
<td>38</td>
<td>0.433</td>
<td>0.351</td>
</tr>
</tbody>
</table>

The E-Schoolbag have been used in many subjects, both traditional subjects (Chinese, Math and English) and other subjects, such as science, geography and biology etc. As we can see from Table 2, the promotion effect of E-Schoolbag on mathematics, geography and biology is small, the value of d is between 0.3 and 0.4. The promotion of Chinese and English learning is moderate, the value of d is greater than 0.5. The promotion of science is significant, the value is greater than 0.8. But the effect of science should be interpreted cautiously because the number of ESs is small.

Table 2. E-Schoolbag’s Effect Size on Academic Achievement by Subject

<table>
<thead>
<tr>
<th>subject</th>
<th>n</th>
<th>Effect size and 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>Chinese</td>
<td>6</td>
<td>0.593</td>
</tr>
<tr>
<td>Math</td>
<td>15</td>
<td>0.339</td>
</tr>
<tr>
<td>English</td>
<td>10</td>
<td>0.583</td>
</tr>
<tr>
<td>Science</td>
<td>1</td>
<td>0.882</td>
</tr>
<tr>
<td>Geography</td>
<td>3</td>
<td>0.382</td>
</tr>
<tr>
<td>Biological</td>
<td>3</td>
<td>0.311</td>
</tr>
</tbody>
</table>

Regarding school level (Table 3), the result of E-Schoolbag’ effect for elementary school (0.468) is higher than that for middle school (0.406), and higher then high school (0.312). But from the overall analysis, the E-Schoolbag’ effect in various school levels is still small.

Table 3. E-Schoolbag’s Effect Size on Academic Achievement by School Level

<table>
<thead>
<tr>
<th>School Level</th>
<th>n</th>
<th>Effect size and 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>Elementary</td>
<td>23</td>
<td>0.468</td>
</tr>
<tr>
<td>Middle</td>
<td>11</td>
<td>0.406</td>
</tr>
<tr>
<td>High</td>
<td>4</td>
<td>0.312</td>
</tr>
</tbody>
</table>
4. Discussion

According to the results of the overall analysis, the overall, average effect size of E-Schoolbag on students’ academic achievement is 0.433. Although the value is a small-to-medium effect size, it still indicates that E-Schoolbag have a positive effect on student learning to some extent. And Compared with traditional paper textbooks, when teacher integrate the E-Schoolbag into the classroom teaching process, it can improve students’ learning interest, learning engagement and learning efficiency (Xu et al., 2013). The result also shows that the effect of E-Schoolbag is vary from the subject. It means that when teachers using E-Schoolbag in their classroom, they need to take full account the differences in subjects. Teachers also should be based on specific application needs to make some changes. Rather than take a fixed E-Schoolbag teaching model, teachers should combine with the characteristics of the subject to design their teaching. As for school levels, compared to high school stage, the application of E-Schoolbag in the primary and secondary stage have a better effect. We can explain this phenomenon based on the current state of education. In high school, students’ learning relatively fixed, and primary and secondary school learning is relatively free (Zhang, Jiang & Lu, 2016). Therefore, This E-Schoolbag-based ubiquitous learning approach may be more suitable for primary and secondary school students, and students may have more curiosity and interest.

5. Conclusions and Limitations

In this study, we use meta-analysis to survey the effect of E-Schoolbag in Chinese schools, and try to derive an exact value to quantify its effect. The findings indicate that E-Schoolbag contribute to students’ academic achievement although they are not as effective as expected. It means that the integration of E-Schoolbag into classroom still a long way to go, and needs to be proved in longer viewpoint. Though government support is very important for the integration of E-Schoolbag in the classroom, schools and teachers should also actively explore the effective way to use E-Schoolbag. Although This study gives an overall, average effect size of E-Schoolbag on students’ academic achievement by using a meta-analysis method, there still have some limitations. The data of Meta-analysis is little, so the results can only explain some of the problems. Future needs to have more research be conducted to compare the effectiveness of E-Schoolbag and the tradition paper textbook.

References

Real-time Analysis of Digital Textbooks: What Keywords Make Lecture Difficult?

Kousuke Mouri*, Atsushi Shimada¹, Chengjiu Yin², Uosaki Noriko³, Vachirawit Tengchaisri⁴ & Keiichi Kaneko⁵

¹, ⁴ Institute of Engineering, Tokyo University of Agriculture and Technology, Japan
² Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
³ Department of Information Science and Technology, Kobe University, Japan
⁴ Center for International Education and Exchange, Osaka University, Japan
⁵ Faculty of Information and Communication, Mahidol University, Thailand

*mourikousuke@go.tuat.ac.jp

Abstract: This paper describes a real-time learning analytics to find learning contents or keywords that students don’t understand in digital textbooks. We developed a digital textbook viewer system that can collect students’ learning logs. By analyzing and visualizing the collected learning logs in real time, teachers can visually find the keywords that students don’t understand during a class. This paper describes the contribution of real-time learning analytics for supporting teachers.

Keywords: learning analytics, digital textbook, real-time analysis

1. Introduction

University classes in Japan are generally conducted in a traditional lecture style. Its advantage is that teachers can convey knowledge to their students precisely. However, it is difficult for them to find keywords that students don’t understand during the class unless teacher asks them. We believe that real-time collection and analysis of students’ learning logs can facilitate teachers to find keywords that their students don’t understand during class.

This study proposed a digital textbook viewer system called AETEL to collect the learning logs (Kiyota et al., 2015, 2016). In order to find keywords that students don’t understand in the digital textbook, this study develops a real-time learning analytics system. By detecting the keywords in real time, the system immediately provides teachers with the keywords that students don’t understand, and teachers are able to explain more in details based on the analysis results.

The rest of this paper is constructed as follows. Section 2 describes our developed digital textbook viewer system to collect learning logs. Section 3 describes our proposed real-time learning analytics for finding keywords that students don’t understand in the digital textbook. Section 4 describes our conclusion and future works.

2. Digital Textbook Viewer System

Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, all the textbooks for elementary, middle, and high schools into digital textbooks (Yin et al., 2014, Ogata et al., 2015). In many countries, the policies of digital textbooks only focus on introducing the technology of digital textbooks into K12 schools (Fang et al, 2012, Shin, 2012). However, little attention has been paid to analyzing the learning logs collected by digital textbooks, although it is important to investigate how these logs can be used to improve digital textbooks and the quality of learning and education (Mouri et al., 2016). In order to find the keywords that students do not understand in the digital textbooks, this study developed a digital textbook viewer system.

Figure 1 shows the interface of our developed digital textbook viewer system. For example, in order to go to the next page, the learner will click “NEXT” button, and this action will be saved as...
“NEXT”. If the learner highlights some row in the digital textbook, he/she will click “Marker” button, and the action will be saved as “ADD MARKER”. For example, when a learner doesn’t understand a keyword in the digital textbook or a learner wants the teacher to give more explanation about it, the text information of the keyword can be collected by using the function.

By using the function, the study can collect keywords that students don’t understand in the digital textbook during class. In the next section, this paper describes the real-time learning analytics for finding keywords that students do not understand in the digital textbook.

![Digital textbook viewer system](Figure 1)

**Figure 1.** Digital textbook viewer system

### 3. Overview of our proposal

Figure 2 (Right) shows the real-time heat-map.

![Real-time heat-map](Figure 2)

**Figure 2.** Overview of our proposed real-time learning analytics
The horizontal axis represents the page number and the vertical axis represents the keywords. The color of each cell represents the number of students who marked the keywords that they did not understand. In this case, teachers see that 11 students can’t understand the keywords, “Call by reference” in real time during the class. As the result, teachers can explain more about it as shown in Figure 2(left).

4. Conclusion and Future Work

This paper introduced a digital textbook viewer system to collect keywords that students do not understand. By using the system, we can collect learning logs to the server, and analyze and visualize them for improving educational quality. Moreover, this paper described a real-time learning analytics for finding keywords that students do not understand in the digital textbook. However, its evaluation is yet to be conducted to examine whether our proposal is useful for teachers, e.g., “whether it hinders their lectures during the class by providing the analysis and visualization results”. The real-time feedback might disturb their lectures if the information load is too much.

With these considerations in mind, we are planning to implement the real-time learning analytics approach in our universities and evaluate it. Also, we will extend our research based on our recent outcome of learning analytics (Mouri et al., 2015; Shimada et al., 2017) and other domains such as career support and language learning (Uosaki et al., 2015) in the future work. We believe that providing teachers with valuable information will lead to improve the quality of education.

Acknowledgements

The part of this research work was supported by the Grant in-Aid for Scientific Research No.16H03078 and No. 17K12947 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.

References

Designing an Intervention for Novice Programmers Based on Meaningful Gamification: an Expert Evaluation

Jenilyn AGAPITO* & Ma. Mercedes RODRIGO
Ateneo de Manila University, Philippines
*jen.agapito@gmail.com

Abstract: Gamification is defined as the addition of game-like elements and mechanics to non-game contexts to encourage certain desired behaviors. It is becoming a popular classroom intervention used in computer science instruction, including CS1, the first course computer science students take. It is being operationalized to enhance students' learning experience and achievement. However, existing studies have mostly implemented reward-based game elements which have resulted to contrasting behaviors among the students. Meaningful gamification, characterized as the use of game design elements to encourage users build internal motivation to behave in a certain way, is contended to be a more effective approach. The foundation of this concept is the 'Self-Determination Theory', which states that there are three components associated with intrinsic motivation: mastery, autonomy, and relatedness. This paper describes the first part of a research on the exploration of how a tool founded on meaningful gamification will affect the achievement and learning experience of novice programmers. It focuses on the design and implementation of a programming-based activity management system embedded with game design elements that map to the different components of the Self-Determination Theory. The elements employed are: feedback cycles, freedom to fail, and progress to support mastery; control to enable autonomy; and collaboration for relatedness. CS1 instructors invited for an expert evaluation generally agree that said elements are present in the system.

Keywords: gamification, novice programmers, CS1

1. Introduction

Computer science educators and researchers are concerned about student retention and attrition (Beaubouef & Mason, 2005; Dehnadi & Bornat, 2006; Hoda & Andreae, 2014; Robins, 2010; Wortman & Rheingans, 2007). CS1, the first course that computer science students take, has a failure rate of at least 30% across institutions (Bennedsen & Caspersen, 2007; Watson & Li, 2014). Students typically enroll in CS1 with little to no background in programming. They are introduced to a fundamental but difficult skill area. This experience shapes their impression of both the course and degree program. However, some researchers speculate that traditional teaching methodologies may not be the ideal choice for programming classes (Hoda & Andreae, 2014, Watson & Li, 2014). In particular, it may no longer be effective for students of this generation. Young people these days are attuned to the use of technologies to reinvent social living, communication, and learning, causing their divergent perspective on the dynamics and operations of the world, including the classroom (Klopfer et al., 2009).

An approach to instruction reinvention dubbed as gamification has recently become popular. Gamification is defined as adding game-like elements and mechanics to a learning process (Deterding et al., 2011; Glover, 2013). It intends to provide users with a more gameful experience and to encourage certain desired behaviors (Deterding et al., 2013) by adding motivational affordances in an environment.

1.1 Gamification in Education

Gamification has become evident in education and classroom design. It is shown to be an effective
method to help at-risk students succeed (Ross, 2011). Several studies focus on gamifying Computer Science subjects (Gibbons, 2013; Ibáñez et al., 2014; Neve et al., 2014; Pirker et al., 2014), including CS1 (Harrington, 2016; Kumar & Khurana, 2012; Sprint & Cook, 2015). They experiment with incorporating game design elements such as badges/reward systems (Harrington, 2016; Ibáñez et al., 2014; Neve et al., 2014; Pirker et al., 2014; Sprint & Cook, 2015), leaderboards (Ibáñez et al., 2014; Neve et al., 2014; Pirker et al., 2014; Sprint & Cook, 2015), point systems and leveling (Ibáñez et al., 2014), and microworlds (Neve et al., 2014) into certain aspects of the learning environment (e.g. home works and practical sessions). Students express preference of a gamified strategy over the traditional because it is able to address their need for fun, pleasure, and cooperation (Kim, 2013). It taps directly into their fundamental desire for recognition, reward, status, competition, collaboration, and self-expression (González & Area, 2013).

1.2 Pitfalls of Gamification

Learning environments are commonly gamified by using badges, levels/leaderboards, achievements, and points because they are relatively easy to implement. This practice is referred to as reward-based gamification (Nicholson, 2015). However, this is effective in contexts that call for short-term change in behavior and those that can continue to supply the rewards (Nicholson, 2015). In the classroom, studies report on disparate reactions from students. Some perceived the method enthusiastically and performed well, whereas others felt demotivated and disengaged. Badges can be perceived as an unattractive alternative to grades (Pirker et al., 2014). Competition-based leaderboards may stimulate only a few and induce a feeling of embarrassment when available for public viewing (Kapp, 2012). Some students continue to fight for the top rank while some disengage altogether (Nicholson, 2013). They generate uneasiness leading students to dismiss it as a good representation of knowledge (DomíNguez et al., 2013). Competition is not fun for some students. Another drawback is when they become too driven by the competition that they lose sight of the real purpose, which is to learn (Sprint & Cook, 2015).

1.3 Meaningful Gamification

Nicholson broadens (Deterding et al., 2011)’s definition of gamification as the use of game design elements to help a user build intrinsic motivation – motivation not triggered by external rewards – to encourage engagement in a specific context (Nicholson, 2015). This is referred to as Meaningful Gamification. The theory behind is known as the Self-Determination Theory (Deci & Ryan, 2002) which states that there are three components associated with intrinsic motivation: mastery, autonomy, and relatedness. Mastery is when one learns to the point of competence; autonomy means having a choice and control of paths; and relatedness is about one’s social engagement. There are contentions that gamifying the classroom entails rigorous research and design efforts to maximize its benefits. Otherwise, it will become ineffective because of the potential adverse effects to some students.

This paper presents the first part of a study to explore the effectiveness of gamification in an introductory programming class and its impact on the learning experience and achievement of novice programmers. It focuses on the design and implementation of a programming-based activity management system embedded with game design elements based on meaningful gamification. The results of an expert review of the system are summarized to determine whether it implements certain game design elements.

2. Research Objective

The main objective of the study is to experiment on the use of gamification in an introductory programming class. It seeks to explore how certain game design elements will impact the learning environment and whether or not it will positively affect the students’ achievement and learning experience. To realize this, the first goal of the study is to design and build a pedagogical tool embedded with game design elements based on meaningful gamification. Prior to actual system development, gamification techniques and methods were evaluated to support the choice of elements and how they
were incorporated into the system. Expert evaluation with CS1 instructors was conducted to assess whether these elements are found in the system.

3. Significance of the Study

Computer science educators and researchers continue to be concerned about student retention and attrition. This research will contribute to the on-going exploration of gamification to increase novice programmers’ motivation. Since most experiments apply gamification through reward-based elements and have been shown to be less advantageous to some, this study delves into a more student-centric design that is anticipated to target their internal motivation to learn. Should the results reveal a positive impact on their achievement and learning experience, game elements that are more effectively applied in learning contexts may be identified. This can possibly spearhead future design efforts in fabricating new ways of delivering CS1 and instrumenting learning tools appropriate for novice programmers.

4. Methodology

This section discusses the design and implementation of a gamified programming-based activity management system for introductory programming classes. It is a web-based platform that allows teachers to manage assessment activities such as quizzes and laboratory exercises typical of CS1. The questions that may be included in the activities are divided into three categories as in (Orji et al., 2013): 1) factual or conceptual; 2) comprehension; and 3) generation questions. Table 1 lists the specific assessment activities the system provides for each category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Assessment Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factual or Conceptual</td>
<td>Multiple Choice Questions, True or False, Identification</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Code Tracing</td>
</tr>
<tr>
<td>Generation</td>
<td>Program Writing</td>
</tr>
</tbody>
</table>

*Conceptual* questions test students’ knowledge of programming elements. Questions assess whether they are able to recall and recognize concepts learned during lectures. *Comprehension* questions check whether students are able to read code by determining the output or the values of a certain variable. Teachers provide code snippets that students have to read. *Generation* questions are comprised of programming exercises that test the students’ ability to write programs.

The students’ view features a coding area for each programming problem that they can use to write their programs (Figure 2). Compilation can be done through a button click. Compiler outputs, if any, are displayed. Program execution is automatically done when compilation results to no errors. In such a case, program output is displayed.

4.1 Evaluation and Selection of Game Design Elements

As discussed, gamifying learning environments do not always achieve the intended outcome. Ineffective classroom gamification could disengage students and in the most serious case, demean and demoralize to the point of dropping out (Nicholson, 2013). Reward-based practices, though easier to employ, do not maximize the potential of gamification. In this light, the choice of elements used in the system was anchored on meaningful gamification, purposely mapping them to the three components of the Self-Determination Theory. The learning module entitled Practical Guide to Meaningful Gamification discussed different ways that can support or oppose these components. A summary is found in Table 2.
Table 2: How to support or oppose autonomy, mastery, and relatedness.

<table>
<thead>
<tr>
<th></th>
<th>How to Support</th>
<th>How to Oppose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy</td>
<td>Providing students with choices</td>
<td>Mandatory participation</td>
</tr>
<tr>
<td>Mastery</td>
<td>Feedback cycles, allowing students to</td>
<td>Lack of feedback, high stake</td>
</tr>
<tr>
<td></td>
<td>fail and retry</td>
<td>consequences</td>
</tr>
<tr>
<td>Relatedness</td>
<td>Teams, competition</td>
<td>Working alone</td>
</tr>
</tbody>
</table>

Providing students choices make them feel in control of the paths they can take. Giving them several options for projects to work on or problems to solve affords them the opportunity to work on something they feel they are good at. This is a modest method of giving students a sense of autonomy.

The concept of “practice makes perfect” helps support mastery. When a system permits mistakes with no associated high stakes, students will be more willing to take risks knowing that they can recover. This is identified as a contributing factor to user engagement. When adopted in the classroom, this offers them a positive perspective on failure as a necessary part of learning (Lee & Hammer, 2011). Allowing the weaker students to reattempt failed work imposes a wider avenue for learning, improvement, and self-actualization (Nicholson, 2013). Immediate feedback is likewise important to help them understand the effects of their actions.

Lastly, designing social experiences can make them feel more like a part of the learning community. Involvement in a group plays a part in one’s academic success (Ohno et al., 2013). Group learning can be a valuable component that maximizes learning gains (Pirker et al., 2014).

These concepts shaped the selection of the game design elements to be included in the system which are listed in Table 3. Further details on each one are discussed in the succeeding sub-sections.

Table 3: Game design elements employed in the system.

<table>
<thead>
<tr>
<th>Game Design Element</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy</td>
<td>Control</td>
</tr>
<tr>
<td>Mastery</td>
<td>Feedback cycles, freedom to fail, progress</td>
</tr>
<tr>
<td>Relatedness</td>
<td>Collaboration</td>
</tr>
</tbody>
</table>

4.1.1 Autonomy

The students are afforded a sense of control by giving them the ability to choose which questions to answer (e.g. quizzes), problems to solve (e.g. laboratory exercises), or out-of-class challenges (optional) to undertake. The system allows teachers to include extra questions/problems and specify the number of required items that the students should answer. Students can choose from the set of questions/problems and answer/solve only up to what is required (e.g. solve 3 out of the 5 problems).

![Figure 1. UI changes when a student answers the number of item required.](image-url)
Activity items will be shuffled. Hence, each student will see them in varying orders. To prevent them from answering more than what is required, the system keeps track of the number of items they have answered so far. Questions, by default, will have blue headers. When they have reached the limit, extra items will be disabled. Header colors will turn to gray and will no longer be editable. Should a student want to change items, he/she can click on the “Clear” button beside each question previously answered. This will enable the extra questions, making them available for him/her to answer again. The “Clear” button is initially hidden and is only displayed when an item is answered. Figure 1 shows these changes in the UI.

4.1.2 Mastery

Mastery is supported through feedback cycles and the freedom to fail. Promptly giving feedback will let the students know whether their answers are correct or not. Freedom to fail gives them the chance to reattempt failed work which will optimally lead to mastery.

The feedback mechanism includes (1) automatically showing the score once an activity has been completed, and (2) marks indicating which items are correct or wrong. Once a student submits an activity, the results plus details about the activity are displayed at the top of the page. See Figure 2.

![Figure 2. Results displayed upon activity submission.](image)

A radar chart illustrating how they fared in the activity is also included (more on this below). A “Show Item Feedback” button that a student can switch on or off is available. When switched on, marks for correct and incorrect items are displayed. Headers of correct items are colored green; incorrect items are red, as shown in Figure 3. The student can then reattempt the activity and try to correct previous mistakes. When switched off, only the score and activity summary are shown. Though reattempts are still allowed, the student would not know which items were marked correct or wrong. This design choice was employed to be able to gather data that will help determine whether students are taking advantage of this feedback mechanism through the clicks on the “Show Item Feedback” button.

Still on mastery, progress for various skills is represented as a radar chart to give the students a general picture of how they are moving forward relative to the system-facilitated activities. This is displayed in their “Profile”, as shown in Figure 4, along with other information. The skills that will be tracked within the system are listed in Table 4 with the question types/activities to assess them. Students’ progress for
‘Concepts’, ‘Tracing’, and ‘Programming’ are based on their scores on the corresponding question types/activities in Table 4.

Table 4: Skills represented in the progress chart.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Question / Activity</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of programming elements</td>
<td>Multiple Choice Questions, True or False, Identification</td>
<td>Concepts</td>
</tr>
<tr>
<td>Ability to trace code</td>
<td>Code Tracing</td>
<td>Tracing</td>
</tr>
<tr>
<td>Ability to correct syntax errors while writing a program</td>
<td>Program Writing</td>
<td>Debugging</td>
</tr>
<tr>
<td>Ability to write correct solutions to programming problems</td>
<td></td>
<td>Programming</td>
</tr>
<tr>
<td>Social/collaborative skills</td>
<td>Peer Mentoring</td>
<td>Tutoring</td>
</tr>
</tbody>
</table>

‘Debugging’ characterizes how students struggle with syntax errors while programming and is evaluated using the Error Quotient (EQ) algorithm, a quantification of students’ compilation behaviors by considering the type, location, as well as frequency of encountered syntax errors (Jadud, 2006a; Jadud, 2006b). The EQ is a single number that may range from 0 to 1. An EQ of 1 implies that a student’s successive compilations ended with exactly the same error. Conversely, an EQ of 0 denotes a student is able to fix syntax errors in subsequent compilations. Data for EQ computations are unobtrusively collected while students write and compile their programs.

‘Tutoring’ represents how students collaborate with other students through peer-mentoring. This is discussed more in the next subsection. Aside from the radar chart on a student’s profile which represents their general performance, each activity will likewise have its corresponding chart representative of their performance in that activity. This is shown in Figure 4. It illustrates the student’s score against the activity’s passing mark. In addition, the chart will modestly show their general performance so they are able to see in one glance how they fared in the activity and how it may have affected their overall performance.

4.1.3 Relatedness

Targeting relatedness is the opportunity for collaboration through peer-mentoring. A student may be awarded with “tutor points” by another when the former mentors the latter. They get to decide who among their classmates they seek for help and whether they will reward them or not. They can do this through a ‘Commend’ button that appears beside a student’s name in the class’ view. ‘Tutoring’ is one among the five skills included in the radar chart discussed previously. This is to assess whether employing such a design element encourages students to work together and to help one another.
5. Expert Evaluation

Prior to testing the system with novice programmers, an expert evaluation was conducted to draw feedback from CS1 instructors with regards the implementation of the different game design elements into the activity management system.

5.1 Participants

Seven instructors from Ateneo de Naga University who have handled CS1 classes participated, two females and five males. They were of varying ages with the youngest at 27 and the oldest at 50. CS1 teaching experience ranged from one to five years. Five of them were familiar with the term “gamification” and two of them have implemented it by conducting a competition-based programming activity in their CS1 classes. Only one (1) used a gamified system, the Programming Contest Control System (PC²) which was developed in support of the ACM computer programming contests.

5.2 Methods

The evaluations began with the instructors answering a demographics questionnaire. An introduction followed which aimed to put context to the evaluation being conducted. Gamification was briefly described particularly because two of the respondents were not familiar with the term. The details and objectives of the project were discussed. This included emphasis on the different game design elements and how they were implemented into the system. Then, a demo of the system was presented. After which, the instructors were asked to answer the debriefing questionnaire.

5.3 The Debriefing Questionnaire

The debriefing questionnaire was a modified version of the gamification inventory, an instrument for the structured assessment of gamification in a given system (Broer, 2015; Broer & Breiter, 2015). Items relevant to the elements incorporated into the system and that map to the components of the Self-Determination theory were selected. Other items that may be applicable were also included. Out of 38 items, twelve (12) were selected to be part of the debriefing questionnaire – one (1) for autonomy; five (5) for mastery; one (1) for relatedness; and five (5) other questions. Sample items are in Table 5.

<table>
<thead>
<tr>
<th>Component</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy</td>
<td>Does the system give the user a feeling of autonomy? Can the user decide what to do next? (autonomy)</td>
</tr>
<tr>
<td>Mastery</td>
<td>Does the system promote learning / mastery? (learning / mastery)</td>
</tr>
<tr>
<td>Relatedness</td>
<td>Does the system provide the user with a feeling of relatedness to other users? Is there social contact? (relatedness)</td>
</tr>
</tbody>
</table>

It used a 5-point scale with response options from Strongly Disagree (1) to Strongly Agree (5). The respondents were to select Neutral (3) if they think the system potentially exhibited the game design element, but relied on user input for it. They were asked to elaborate on their levels of agreement. If they agree/strongly agree that an element is implemented, they were asked to identify the specific feature/s of the system that exhibits the element.

5.4 Findings

As mentioned, the game design elements employed are: control to enable autonomy; feedback cycles, freedom to fail, and progress to support mastery; and collaboration for relatedness. The expert evaluation was conducted to specifically assess the presence of these elements along with other items that seemed relevant.
5.4.1 Autonomy, Mastery, Relatedness

The frequency of responses for the questions that fall under autonomy, mastery, and relatedness is summarized in Figure 5.

![Figure 5](chart.png)

**Figure 5.** Frequency of responses for autonomy, mastery, and relatedness questions.

The evaluators agree that the system provides for autonomy by allowing students to select items to answer in a certain activity. This gives the students a sense of control over the questions that they would like to answer, allowing them to choose those they feel they are capable of answering.

They generally agree that the elements for mastery are implemented. The system provides students with a sense of accomplishment through the radar chart and instant feedback (Q2). According to them, the radar chart can motivate the students to be an “expert” in all areas or on certain favored areas. It may encourage students to work more. The radar chart, activity statistics, and immediate feedback can give them a feeling of competence (Q3). Displaying the passing score in an activity’s radar chart shows the students that they must achieve certain marks to get better scores. Learning/mastery is promoted by allowing students to re-attempt activities. The repetition and familiarization can enhance learning. Through the feedback mechanism, a student can understand which parts of an activity he/she correctly understood. However, one respondent raised a concern about the possibility of students avoiding certain questions – those that may be too difficult for them (e.g., programming). Meaningful feedback is provided through displaying the scores, activity statistics, and the correct/wrong answers (Q4). Progress is provided through the radar chart and the summary statistics allow students to sense how well they are doing (Q5).

As for relatedness, a couple of respondents agree that awarding tutor points can encourage peer-mentoring among the students. Majority, however, responded with neutrality, as seen in Figure 7. According to them, subjectively giving out points may not necessarily encourage students to mentor their peers and feel more connected with one another. One respondent even disagreed and supposed that students might just “pair-up” and give each other points.

5.4.2 Others

Four out of seven agree that the different question types available in the system can provide for challenging tasks for the students (Q1). However, teachers play an important role in constructing challenging questions. One of them disagreed and suggested that allowing students to view other students’ profiles can help support this game element. Showing the progress of the students through the radar chart may increase the students’ attachment to outcome (Q2). Since the chart

![Figure 6](chart.png)

**Figure 6.** Frequency of responses for other items.
depicts their progress in terms of the different activities facilitated through the system, this can possibly urge them to work to further their performance in certain or all areas. Positive emotions can be triggered by displaying their scores and progress, especially if they see improvements in their scores after several attempts (Q3). Most of them agree that incorporating the different game design elements can spark more interest and increase focus/attention (Q4). Lastly, the radar chart and summary statistics can represent students’ behavior/performance (Q5).

6. Conclusion and Future Works

This paper presented the design, implementation, and expert evaluation of a gamified activity management system for novice programmers as part of a larger study on the exploration of the effects of gamification on students’ achievement and learning experience in CS1.

Gamification is becoming popular as a classroom intervention. However, previous studies mostly applied reward-based game elements and were shown to be not effective for some. Hence, this research intends to explore the impact of a tool founded on meaningful gamification. The selection of game design elements implemented into the system should map to the different components of the Self-Determination Theory. The elements employed are: control to enable autonomy; feedback cycles, freedom to fail, and progress to support mastery; and collaboration for relatedness.

Expert evaluation was conducted to seek insights from CS1 instructors with regards to the presence of said elements in the system. The respondents generally agree that the elements for autonomy and mastery existed. Though, they were more neutral with regards to relatedness. They do not think that subjectively awarding tutor points encourages students to mentor one another. As for the other additional items, responses were generally positive.

Moving forward, the system will be tested on novice programmers. Control and experimental groups will be asked to use a non-gamified and gamified version, respectively, to be able to gather data that can possibly differentiate students’ experience and performance. Data will then be analyzed to determine whether the students’ behaviors in response to the system can essentially be attributed to the game design elements.

Acknowledgements

We would like to thank the instructors for responding to the invitation to evaluate the system. We also like to thank the Department of Computer Science, Ateneo de Naga University for providing a venue for the evaluation. Much gratitude to the Department of Science and Technology–Engineering Research and Development for Technology (DOST-ERDT) for the scholarship awarded to the student researcher. Lastly, we thank the Ateneo Laboratory for the Learning Sciences (ALLS), Ateneo de Manila University for supporting this study.

References


Virtual Currency as Gamification for Learning in a Disaster Museum to Increase the Number of Revisitors

Hiroyuki MITSUHARA* & Masami SHISHIBORI
Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Japan
mituhara@is.tokushima-u.ac.jp

Abstract: Disaster education plays an important role in protecting our lives and information, and communication technology-based disaster education (ICTDE) has received much attention. We focus on virtual currency as gamification to enable visitors to learn about disasters more deeply while having fun at a disaster museum that serves as a central facility for disaster education; accordingly, visitors can be motivated to revisit the museum. In our gamified ICTDE program, “Knowledgeonaire,” visitors can take quizzes while observing exhibits and obtain VC as a reward when they provide correct answers. In addition, visitors will think about how to make effective use of their VC when receiving VC, buying hints, and buying gifts. We developed a web-based Knowledgeonaire system (K-System), which has a simple mechanism of quiz presentation using near field communication (NFC). Through a preliminary comparative experiment, we discovered that K-System did not perform better than a non-Knowledgeonaire system (NFC-based simple quiz presentation system), but can motivate visitors to revisit the museum.

Keywords: Virtual currency, gamification, quiz, learning at museum, disaster education

1. Introduction

Disasters that cause catastrophic damage are increasing in number around the world. In Japan, for example, the Great East Japan Earthquake and the resulting tsunami caused many deaths. Each time a disaster occurs, we are motivated to learn about disasters; thus, disaster education plays an important role in protecting our lives. However, disaster education is yet to be fully established. For example, evacuation drills are often conducted as part of traditional disaster education programs, but participants do not necessarily commit to such drills. We cannot know exactly when and where a disaster is likely to occur and therefore, we often regard the occurrence of a disaster as someone else’s problem. In other words, we are prone to lower the priority of disaster education and do not acquire sufficient knowledge about disasters through conventional disaster education programs.

One approach to overcoming this situation is to diversify disaster education. For diversification, information and communication technology (ICT) has attracted much attention, and there have been various kinds of ICT-based disaster education (ICTDE) programs. For example, programs using digital games enable participants to learn about disasters while having fun through high interactivity and the use of audio-visual effects (e.g., Tsai et al., 2015; Wahyudin & Hasegawa, 2015; Kawai et al., 2016). Programs involving interactive simulations enable participants to plan proper evacuations from visualizations of panoramic disaster damage (e.g., Kobayasi at al., 2008) or by evacuating virtual three-dimensional worlds (e.g., Dunwell et al., 2011; Capuano & King, 2015). In addition, disaster big data, collected from social media (e.g., Twitter) and sensor nodes (e.g., Global Positioning System on smartphones), clarify disaster phenomena and provide lessons for surviving disasters (e.g., Power et al., 2016).

In this study, we deal with ICTDE in a disaster museum that serves as a central facility for disaster education in Tokushima prefecture (approximately 0.78 million citizens) in Japan. It is...
expected that Tokushima prefecture will be damaged by a massive earthquake and tsunami in the near future; thus, Tokushima citizens have an interest in disasters, and many of them must have visited the museum at least once (e.g., as an extracurricular lesson for students). The museum has experience-oriented simulators (e.g., earthquakes, storms, and smoke) and permanent exhibits (e.g., pictures, video materials, tangible models, and expository panels). In addition, the museum regularly conducts special exhibits (e.g., reports on the latest disaster cases), and museum attendants offer visitors supplemental explanations regarding the exhibits. However, the museum has difficulty in attracting revisitors even though it is recognized as an important facility for learning about disasters, because disasters contain serious, negative themes (e.g., death). In other words, the citizens look away from disasters and are not motivated to revisit the museum. We believe that repeated revisits will lead to acquiring broad and diverse knowledge regarding disasters.

To increase the number of revisitors, we focus on gamification and introduce virtual currency (VC) to digital quizzes on permanent and special exhibits. We refer to this gamified ICTDE program as “Knowledgeonaire.” In Knowledgeonaire, visitors can take quizzes while observing exhibits and obtain VC as a reward when they provide correct answers.

The remainder of this paper is organized as follows. Section 2 describes the fundamental idea and the developed system of Knowledgeonaire. Section 3 reports a preliminary experiment. Section 4 summarizes this study.

2. Knowledgeonaire

VC can be regarded as extrinsic motivation. In Knowledgeonaire, visitors can learn about disasters to earn VC. Some people claim that visitors should learn about disasters based on intrinsic motivation. However, citizens who do not have or retain an interest in disasters will not revisit the museum. We believe that VC provides a strong incentive that can motivate such citizens to revisit the museum.

2.1 Gamification and Virtual Currency (VS) in Education

Gamification is typically defined as “the use of game design elements in non-game contexts” (Deterding et al., 2011) and “a process of enhancing a service with affordances for gameful experiences in order to support user’s overall value creation.” (Huotari and Hamari, 2012) The game design elements, which encourage participants to struggle with tasks with continuous motivation (affordances), are provided as such rewards as points, badges, and leaderboards. Gamification has attracted attention in various domains, including education (Seabirn and Fels, 2015). In the domain of education, gamification is a promising approach to increasing learning motivation, engagement, and effectiveness (Erenli, 2013; Dicheva et al., 2015; Lister, 2015). On the other hand, Hanus and Fox (2014) reported that gamification was not effective in motivation, satisfaction, and empowerment owing to such conditions as the degree of learners’ intrinsic motivation.

There is no single definition of VC. For example, VC is defined as “a type of unregulated, digital money, which is issued and usually controlled by its developers, and used and accepted among the members of a specific virtual community.” (European Central Bank, 2012) Another concise definition indicates that VC is used only in closed worlds (e.g., on-line games). Some studies have focused on the kind of VC that can play a positive role in gamified learning. For example, O’Donovan at al. (2013) introduced VC in lecture-related digital quizzes and puzzles enabling university students to buy additional answering chances and hints. Other types of VC allow students to extend a due date for homework submission (Goehle, 2013) and buy a passing grade and their final course grade (Chen et al., 2015).
2.2 Digital Quiz

As the platform for Knowledgeonaire, we focus on digital quizzes about exhibits. Digital quizzes have often been introduced into game-based learning in museums. For example, Sung et al. (2013) developed a museum guidebook system that presents quiz questions about exhibits on a tablet computer and enables visitors to interact with virtual characters in game-style materials. Xhembulla et al. (2014) developed a mobile learning system that provides visitors with mini-games including quizzes about exhibits and allows visitors to explore the museum freely according to their interests and agendas. Mikalef et al. (2013) reported that interactive quizzes with a time limit immersed visitors in exploratory learning and heightened their performance in the post-assessment test.

In Knowledgeonaire, digital quizzes (multiple-choice questions) are presented sequentially to prompt visitors to explore exhibits in a fixed order. Visitors might dislike the sequential presentation, because they cannot answer quizzes according to their interests and agendas; it might differ from the predominant learning styles in museums (i.e., free-choice learning and self-directed learning). In terms of learning effects, however, we think that visitors should learn about disasters while considering the context (e.g., earthquake, tsunami, and then aftershock). In terms of learning motivation, we think that a sequential presentation can achieve story-based gamified learning (e.g., a mystery-solving game that suggests which exhibit has the next question). Museums occasionally aim at increasing learning motivation and effectiveness by adopting an approach different from the usual. The typical approach is to adopt controlled learning styles (e.g., orienteering). We believe that the sequential presentation will enable the visitors to learn about disasters more deeply while having fun.

The sequential presentation, which designates the first and last quizzes, can be used to recognize visitors’ learning levels, i.e., how often they have revisited the museum, by counting the number of times they reach the last quiz. Our digital quiz can encourage visitors to revisit the museum by changing quiz questions according to their learning levels; a revisitor is given higher-level quiz questions as the number of his or her revisits increases. This stepwise presentation enables revisitors to acquire broad and diverse knowledge regarding each exhibit.

2.3 Virtual Currency Framework

In Knowledgeonaire, visitors can acquire a fixed amount of VC each time they provide the correct answer on a quiz. On the other hand, they lose a fixed amount of VC when they provide an incorrect answer—their minimum amount of VC is zero. The VC’s monetary aspects are (1) visitors can buy hints for quizzes, and (2) they can buy real commemorative gifts of the museum. This VC framework encourages visitors to learn strategically, i.e., to think about how to make effective use of their VC. Visitors make the following decisions about their VC (Figure 1).
2.3.1 Receiving VC

VC is accumulated as a donation from visitors who earned their VC by having correct answers. At the beginning of Knowledgeonaire (every visit to the museum), visitors can decide whether or not to receive an amount of VC from the donation bank. The donees (i.e., the visitors who receive VC) have the following duties: (1) they must return the full amount of their received VC to the donation bank after the last quiz and (2) they must donate half of their earned VC (reward received for providing correct answers) to the bank. Visitors who need hints can receive the donated VC.

2.3.2 Buying Hints

For each quiz, visitors who have a specific amount of VC can buy a hint with it. The hint is a short text that narrows down the correct answer from several options. Visitors who do not want to lose their VC will observe exhibits carefully and guess the correct answers without hints.

2.3.3 Buying Gifts

After the last quiz, the visitors can decide whether or not to buy the gifts with their earned VC. Note that before buying the gifts, the donees must fulfill their obligation regarding the donation (as described in 2.3.1)—they can decide to buy gifts with their remaining VC.

2.4 Knowledgeonaire System

We developed a web-based Knowledgeonaire system (K-System), which has a simple mechanism of quiz presentation using near field communication (NFC); the quizzes are presented on a standard web browser installed in NFC-enabled mobile devices (e.g., smartphones and tablets). Figure 2 schematically shows the composition of K-System. Each major exhibit has an NFC tag with the corresponding identifier (ID). Each exhibit’s ID corresponds to a quiz (a multiple-choice question, options, and the correct answer). The web server has a common data interface and a database that consists of exhibit, quiz, visitor, log, gift, and other tables. When a visitor reads an NFC tag using his or her mobile device, the web server receives the tag’s ID from the device and sends the corresponding question to the device. When the visitor answers the presented question, the web server receives and judges his or her answer and sends the feedback depending on whether it is correct or incorrect.

K-System’s user interface is optimized for a web browser in the vertical position of a mobile device (Figure 3). To begin Knowledgeonaire, a visitor reads the initial NFC tag at the reception desk; at the first visit, the visitor is required to complete visitor registration (i.e.,
create his or her login account for K-System). After the successful login, the visitor moves to a page for receiving VC and determines the amount of VC to be received from the donation bank. He or she searches for the first quiz tag (exhibit) according to a receptionist’s instruction. Immediately after the visitor reads the proper tag, the corresponding quiz consisting of a short text and several options is presented. If necessary, the visitor can buy a hint with his or her VC. Immediately after the visitor answers the quiz (i.e., selects one of the options), feedback is presented that includes whether the answer was correct or incorrect, the current amount of the visitor’s VC, expository text, and indication, or implication of the next tag. The visitor reads the final tag at the reception desk, and a shopping page is presented that enables him or her to buy gifts from the list (name, price, and picture).

![Figure 2. Composition of Knowledgeonaire System]

3. Preliminary Experiment

To examine how Knowledgeonaire influences visitors, we conducted a preliminary comparative experiment at the museum using two types of systems: K-System and a non-Knowledgeonaire quiz system (NK-System) that used NFC-based simple quiz presentation but not VC.

3.1 Settings

Thirty-one visitors (7 to 50 years old) participated in this experiment and were divided into two groups: Group-K (14 participants who used K-System) and Group-NK (17 participants who used NK-System). They were instructed on how to use their assigned systems and began taking
quizzes (Figure 4). We prepared eight quizzes for eight exhibits and made the following settings for VC, referred to as “TOK” (currency unit)—if participants of Group-K do not receive VC from the donation bank and answer all of the quizzes correctly without hints, they earn TOK800.

```plaintext
# At Beginning
if I receive VC then
  MyReceivedVC = x;
  MyVC += x;
endif

# When Answering Quiz
if I buy hint then
  MyVC -= 50;
endif
if My answer is correct then
  MyVC += 100;
else
  MyVC -= 50;
end if

# After Last Quiz
if I received VC then
  MyEarnedVC = MyVC–MyReceivedVC;
  MyVC = MyEarnedVC * 0.5;
  DonationBank += (MyReceivedVC + MyVC);
end if
while MyVC > 0 do
  if I buy gift then
    I select gift(i).
    if MyVC > price[i] then
      MyVC -= price[i];
      I receive gift(i).
    end if
  else
    exit;
  end if
end while
```

After the final quiz, the participants of Group-K can buy gifts (e.g., a water pistol for TOK650). The participants of both groups answered a questionnaire regarding the systems they used (five-point Likert scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). The questionnaire consisted of the following questions:

- **Q1.** This quiz system enhanced my interest in disasters.
- **Q2.** This quiz system enhanced my knowledge of disasters.
- **Q3.** This quiz system was easy to use.
- **Q4.** This quiz system motivated me to revisit this museum.
- **Q5.** Virtual currency of this quiz system motivated me to answer quizzes. [only for Group-K]

After the final quiz, the participants of Group-K can buy gifts (e.g., a water pistol for TOK650). The participants of both groups answered a questionnaire regarding the systems they used (five-point Likert scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). The questionnaire consisted of the following questions:

- **Q1.** This quiz system enhanced my interest in disasters.
- **Q2.** This quiz system enhanced my knowledge of disasters.
- **Q3.** This quiz system was easy to use.
- **Q4.** This quiz system motivated me to revisit this museum.
- **Q5.** Virtual currency of this quiz system motivated me to answer quizzes. [only for Group-K]

### Figure 4. Snapshots of Preliminary Experiment

#### 3.2 Results

Nine (64.2%) of the 14 participants of Group-K completed all of the quizzes, i.e., the other participants stopped taking quizzes in the middle. On the other hand, twelve (70.5%) of the 17 participants of Group-NK completed all of the quizzes. The mean values of the learning time (i.e., time between the first and last quizzes) of Group-K and Group-NK were 1,009 and 991
seconds, respectively—the incomplete participants were eliminated from the calculation. A two-tailed t-test showed that there was no significant difference between the two groups. Finally, five participants of Group-K had to return all of their VC to the donation bank. The mean amount of the other participants was approximately tok280, and a few participants were able to buy the less expensive gifts using their VC. We did not find participants of either system who revisited the museum within one week.

The mean values of the correct answer rates of Group-K and Group-NK were 0.18 and 0.28, respectively. A two-tailed t-test showed that there was a significant difference at the 1% level between the two groups.

Table 1 shows the mean values, standard deviations, p-values (two-tailed t-test) for the questionnaire results. For all of the common questions (Q1 to Q4), the mean values of Group-K were lower than those of Group-NK. For Q1 and Q4, the mean values showed significant differences at the 5% level. For Q2, the mean values showed a significant difference at the 1% level. For Q3, there was no significant difference. The mean value of Q5 was favorable. We obtained just one remark on VC from free descriptions: “It is a good idea that VC is used to motivate taking quizzes.”

Table 1: Results of questionnaire (* p < 0.05, ** p < 0.01)

<table>
<thead>
<tr>
<th></th>
<th>Group-K Mean (SD)</th>
<th>Group-NK Mean (SD)</th>
<th>t-test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>4.21 (0.80)</td>
<td>4.82 (0.39)</td>
<td>0.018 *</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>4.28 (0.72)</td>
<td>4.94 (0.24)</td>
<td>0.005 **</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>3.78 (1.18)</td>
<td>4.11 (1.05)</td>
<td>0.417</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>4.28 (0.82)</td>
<td>4.82 (0.39)</td>
<td>0.038 *</td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>4.42 (0.85)</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

3.3 Consideration

Concerning the completion rate and learning time, we think that participants of both groups took quizzes similarly. The completion rates of both groups were lower than we expected. It is possible that participants had compelling reasons to finish quickly (e.g., tight schedule) and/or they were frustrated with the difficult quizzes; actually the correct answer rates of both groups were not high. Although the completion rate of Group-NK was higher than that of Group-K, two participants of Group-NK completed all of the quizzes in an extremely short time (419 and 552 seconds); concerning the learning time, therefore, the standard deviation of Group-K (166) was smaller than that of Group-NK (315). These two participants might not have learned deeply by observing the exhibits, because the simple quiz presentation did not provide the participants with any disadvantages, even when they answered incorrectly. We supposed that participants of Group-K would learn deeply to earn their VC (or not to lose their VC) and accordingly, spend more time for learning. However, the comparative results of the learning time and the correct answer rates might indicate that participants of Group-K did not learn deeply. From these unexpected results, we must conclude that K-System cannot always encourage visitors to learn deeply at the museum.

For both groups, the mean values of all of the questions were favorable and indicate that visitors can positively accept both systems and be motivated to revisit the museum. This indication might be supported by the mean values of Q4. However, we expected that the mean values of Group-K would be higher than those of Group-NK. One conceivable reason why K-System did not dominate over NK-System is that participants of Group-K did not really find fun in the VC framework. In other words, they did not care about or had difficulty in understanding how to make effective use of their VC. On the other hand, participants of Group-NK might have concentrated on taking quizzes. We think that the simple quiz
presentation intrinsically enabled visitors to learn while having fun at the museum. If we interpret these questionnaire results in a negative light, K-System does not necessarily predominate, and VC is to be regarded as just an option.

Free descriptions included some positive remarks regarding the NFC-based simple quiz presentation: “Learning disasters by taking quizzes is fun, especially for children,” “Quizzes prompted me to observe exhibits carefully,” “I learned about disasters effectively by answering quizzes,” and “Quizzes made me interested in exhibits.” However, the mean values of Q3 were lower than those of other questions. In other words, the usability of both systems was low. We think that the low usability resulted from the low sensitivity of the tablets’ NFC readers used in this experiment. For high sensitivity, we must prepare a high-sensitivity NFC reader and take account of NFC tags’ attachment positions, because sensitivity occasionally depends on surrounding environments of the tags.

In this experiment, we could not control the experimental conditions; for example, visitors to the museum were diverse, and the participants were not homogeneous between the two groups in terms of such factors as age and estimated staying time. Therefore, the experimental results appear to lack persuasiveness in the comparison. In addition, we have not examined in the longer term whether the participants actually revisited the museum after this experiment. These limitations must be removed in a subsequent larger-scale experiment.

4. Conclusion

To increase the number of revisitors at a disaster museum, K-System uses NFC-based quiz presentation and VC. K-System encourages visitors to learn about disasters (exhibits) deeply while having fun and accordingly motivates them to revisit the museum; visitors can take quizzes while observing exhibits and obtain VC as a reward when they provide the correct answers. In addition, visitors can be expected to think about how to make effective use of their VC when receiving VC, buying hints, and buying gifts. Questionnaire results obtained from a preliminary comparative experiment indicated that visitors can positively accept K-System and be motivated to revisit the museum. At the same time, however, overall experimental results indicated that K-System cannot always realize its aim. To increase the number of revisitors, we must pursue a simplified VC (i.e., gamification) that balances fun and usability (e.g., additional operations).

In the future, we must reconsider the concept of K-System and improve it from the refined concept. If it is too complicated for visitors, the VC framework (e.g., donation) must be simplified or minimized. If visitors feel uneasy about security issues on VC (e.g., hacking and improper trade), the VC framework (e.g., user authentication) must guarantee the security. Furthermore, we must evaluate K-System in more detail through a more controlled larger-scale experiment while increasing the number of quizzes at the museum.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research (C) No. 15K01026 from the Japan Society for the Promotion of Science. We would like to thank Nobusada, T. and Tokushima Prefectural Disaster Prevention Center.
References


“Go Kahoot!” Enriching Classroom Engagement, Motivation and Learning Experience with Games

Sherlock A. LICORISH*, Jade Li GEORGE*, Helen. E. OWENb & Ben DANIELc

aDepartment of Information Science, University of Otago, Dunedin, New Zealand
bDepartment of Psychology, University of Otago, Dunedin, New Zealand
cHigher Education Development Centre, University of Otago, Dunedin, New Zealand
*sherlock.licorish@otago.ac.nz

Abstract: Technology is being increasingly integrated as a part of teaching in view of enhancing students’ engagement and motivation. Game-based student response systems in particular can motivate engagement, and ultimately, improve students’ learning experience. In this paper we report on the outcomes of employing a game-based student response system, Kahoot!, in an Information Systems Strategy and Governance course at a research-intensive teaching university in New Zealand. In order to examine the efficacy of the system in engaging students during lectures, we conducted semi-structured interviews with students to learn about the extent to which Kahoot! contributed to better engagement and enhanced learning experience. We also explored students’ views about Kahoot!’s influence on classroom dynamics, motivation and the learning process. Overall findings reveal that the deployment of Kahoot! enriches the quality of student learning in the classroom, with the highest influence reported on classroom dynamics, engagement, motivation and improved learning experience. We also learned that the use of games in the classroom can largely minimise distracting classroom behaviours and activities, and improve the quality of teaching and learning beyond what is provided in conventional classrooms (e.g., normal PowerPoint slides and chalk and talk).

Keywords: Game-based Student Response Systems, Kahoot!, Classroom Dynamics, Engagement, Motivation, Learning

1. Background and Research Questions

With mobile devices being ubiquitous and Bring Your Own Device (BYOD) being almost widespread in contemporary classroom teaching, technology is increasingly being integrated into many aspects of classrooms to facilitate assessments, enhance student engagement, motivation and learning. Technology is also frequently used to render difficult topics more engaging and interesting (Prensky, 2001). In supporting such efforts, there has been a shift from student response systems (SRS) such as “clickers” and “zappers” to more contemporary game-based student response systems (GSRS) such as Kahoot! and Socrative systems (Wang, 2015). GSRS are game-based systems where the teacher designs interactive quizzes projected as regular lecture slides to enable students to respond to questions using a web browser on their digital devices in a game-show like situation. Quizzes may be enhanced with images and videos, and the teacher is able to control the pace of play. Students are awarded points for answering questions correctly, and the timeliness of correct responses also impacts the points awarded. Students’ points are displayed on the screen as in game shows, and this drives students to get to the top of the leader board.

The utilisation of games to support engagement in a range of academic settings, including for learning Computer Science, Mathematics and Physics was found to enrich student learning experience (El-Nasr & Smith, 2006). Similar to other mobile device teaching tools, GSRS require participants to activate relevant previous knowledge, and assess their performance as they play and learn content of a subject (Papastergiou, 2009). Further, the multi-sensory, experiential nature of games also enhances students’ problem-solving and critical thinking skills (e.g., McFarlane, Sparrowhawk, & Heald,
(2002), as they cannot advance to the “next level” without attempting to answer questions.

Early research shows that integrating GSRS (e.g., Kahoot!) into regular classroom lectures contributes to improvements in student engagement (Wang, 2015). In addition, our initial observations during Information Science lecture sessions over the past two years suggest that such tools excite students to actively engage in lectures and contribute to the learning environment. Furthermore, consistent with previous research, GSRS enhance positive classroom dynamics (Rosas et al., 2003) and improve students’ interactions with their peers and lecturers. Also, GSRS can motivate those students who may not normally participate in class discussion (Wang, 2015). GSRS were also reported to improve overall class attendance (Cardwell, 2007), and lecturers also found GSRS to be useful teaching tools in supporting personalisation of learning (Wang, 2015).

Despite increasing utilisation of GSRS, it remains unclear the extent to which GSRS can improve learning beyond what would be expected from conventional teaching methods. In addition, it is still unknown whether or not GSRS can improve students’ academic performance. For instance, a meta-analysis revealed that, in over half of the studies investigated, students’ performance did not significantly differ between the use of SRS and traditional learning methods (Randel, Morris, Wetzel, & Whitehill, 1992). Educational games are activity-driven learning tools that often require students to complete special “missions” in order to advance their learning. Until recently, educational games have only supported engagement, with limited understanding of their contribution to improving academic performance. Students can at times learn the correct actions and answers through trial and error rather than by actively reflecting on their learning (Kiili, 2005). There is also a body of research suggesting that high performing students are reluctant to use games as legitimate learning tools (Squire, 2005).

On the contrary, Papastergiou (2009) found that games improved students’ knowledge of computer memory systems to a greater extent than other computer-mediated learning tools, namely, educational websites. This may be due to the experiential nature of GSRS, which likely enhances learning for (at least) two main reasons. Firstly, GSRS induce a “flow” experience where the students become completely absorbed in the goal-driven task expected of them (Kiili, 2005). In other words, students’ experiences “flow” when the games’ interfaces are user-friendly and do not detract attention from the task, and when their abilities match the skill level of the game (Kiili, 2005). Secondly, games encourage the students to collect data, test their own knowledge, seek feedback, draw conclusions and make generalisations so that their knowledge is applied to future learning situations (Kiili, 2005).

In addition to enriching learning, the effectiveness of GSRS depends on whether students perceive the games as appealing, accessible, useful and of high quality. Papastergiou (2009) also found that students rated GSRS as more appealing and more valuable as an educational tool compared to other performance-tracking educational websites that contained the same content. In spite of a small “wear off effect” of long-term GSRS use on student communication and enjoyment (Wang, 2015), students who continued to use GSRS throughout the semester reported positive impacts on learning and engagement similar to the excited new users. Students also commented that, even after a whole semester of using GSRS, they were still motivated to do additional study to prepare for the weekly quizzes.

The simple Likert-scale measures (on their own) that were previously used for GSRS evaluations are not necessarily adequate for understanding the complexities in human behaviour, and particularly those related to students’ engagement, motivation and learning (e.g., Ke, 2009). The literature suggests that exploring users’ experience with game-based technology may be better suited to qualitative survey-based approaches rather than quantitative measures (Nacke, Drachen, & Göbel, 2010). Although Wang (2015) utilised GSRS Likert-scale evaluations with students’ open-ended comments, the data were only analysed quantitatively, and thus, it remains unclear whether semi-structured interviews were conducted to generate answers to specific questions, necessitating further exploration of whether students’ perceptions of GSRS remain the same or can change over time. Finally, the Likert scales were also not always consistent with students’ open-ended comments (Wang, 2015). For instance, the positive effect of GSRS on communication reduced over time, but students explained this was because of impending assessments, and they wanted to focus more on the quiz content than communicate with other students. There is thus need for exploratory studies to unpack if and when GSRS tools help, in support of our understanding of classroom dynamics and enhance students’ engagement, motivation and learning.

We broadly conceptualised classroom dynamics as the interaction between students and lecturers. Student engagement relates to the level of attention, curiosity, focus and interest that
students show during the course. Motivation is the persuasion to be engaged and interact in the classroom. Learning is defined as the knowledge and skills that students attain that is directly attributed to their involvement and participation in the course.

Overall our research aims to contribute to the better understanding of accrued benefits of using GSRS in learning, and to gauge the extent to which the use of Kahoot! can enhance students’ learning experience. More specifically, our objective was to understand how students experienced the use of Kahoot!s and to explore the extent to which this interactive technology influences classroom dynamics, engagement, motivation and learning. In our study we addressed the following four research questions:

**RQ1.** What effects does Kahoot! have on classroom dynamics?

**RQ2.** Does the use of Kahoot! influence students’ engagement, and how?

**RQ3.** To what extent does the use of Kahoot! influence students’ motivation towards learning?

**RQ4.** How does the use of Kahoot! enrich learning according to students’ experiences?

### 2. Methods and Procedures

We employ a qualitative approach to address the four stated questions. We believe that a qualitative research approach is relevant to utilise in this study because the phenomenon being studied is not easily distinguished from the context in which it is observed (Yin, 2013). Using an explorative case study, we intend to unravel complex perceptions and issues relating to the use of Kahoot! in the context of students’ engagement, motivation and learning.

The game-based student response system (Kahoot!) was used as a part of a third year course on Information Systems Strategy and Governance (INFO322) in the second semester of 2016 (between July and November). This tool was used in four (4) different ways during seven (7) different lectures, with a duration of about 30 minutes on average (students could also play Kahoot! outside of the classroom). These include: to quiz students on various topics to understand their competence before tailoring lesson plans, for exploring students’ knowledge of topics after they were delivered in lectures, to help students to validate their comprehension and understanding of topics by having them design their own Kahoot! assessments which were then collectively played, and for fun where the focus was on topics unrelated to the course (e.g., sports). Moreover, the Kahoot! game environment was designed with many interactive features (including suspense music), where students used mobile devices (smartphones, tablets and laptops) to join the games and answer questions, and responses to their choices were visualised (illustrated in Figure 1).

![Image of Kahoot! interface](image)

**Figure 1.** Game show interface projected on screen and on mobile device
2.1. Sampling and Participants

At the end of the course students were interviewed using a semi-structured approach. Purposive non-probability sampling was used to recruit students enrolled in the course. The study was announced and its purpose explained during the final lecture. The study received human and behavioural ethics approval from the university in which the study was conducted.

Fourteen students (10 male, 4 female) agreed to participate in the study. The sample size is deemed adequate for the chosen (purposive) sampling method as the possible pool of participants is already restricted (Marshall, 1996). Students agreeing to participate were asked to spare 20 minutes of their time for the semi-structured interview where they were asked questions relating to the use of “Kahoot!” during the course (interviews took between 15-20 minutes). The questions were focused on understanding students’ experiences using Kahoot!, and the tool’s influence on classroom dynamics, their engagement, motivation and learning. Students were also asked to give suggestions for alternative uses of “Kahoot!”, and describe their general experience with the tool. Sample questions included: “How do you feel about the changes in the INFO322 classroom dynamics brought about by Kahoot!” and “Do you feel that Kahoot! increase/decrease your engagement during the INFO322 course, and how did it increase/decrease?”.

2.2. Data Processing and Analysis

Students’ responses to the interviews were transcribed by the second author, i.e., verbatim. The transcripts were identified by author ID, interview time, questions and responses, and students were treated as the units of analysis. Thereafter, our analyses of the content were performed.

We adopted an inductive (bottom-up) approach to content analysis to test whether clear themes relating to classroom dynamics, engagement, motivation and learning appeared in the data (Patton, 1990). The procedure involved open coding where the interviews were read and re-read for familiarisation and initial codes were identified based on explicit, surface-level semantics in the data, rather than implicit responses and preconceptions (see Braun & Clarke, 2006). Through axial coding, codes were recombined and connections were formed between ideas. Then, we used thematic mapping to restructure specific codes into broader themes. Finally, following Braun and Clarke’s (2006) selective coding procedure, the resulting themes were refined and organised into a coherent, internally consistent account, and a narrative (“story”) was developed to accompany each theme. Themes were extracted from answers provided in response to interview questions, which targeted understandings around classroom dynamics, students’ engagement, motivation and learning.

Descriptive statistics were used to summarise participants’ demographics, including the gender distribution (noted above), ages, years of study, hours spent studying and performance in the course. Performance was measured based on coursework (i.e., case critique, case study and class project) and final exam grades, where students tended to perform better in the former assessment. Our summary statistics show that overall students’ average age was 21.4 (median=21.5 and Std. Dev.=1.3), and they had completed close to 4 years of study (mean=3.4, median=4, and Std. Dev.=0.8).

We observed that students spent around 6 hours a week studying for the course (mean=6.1, median=6.0, and Std. Dev.=2.7), their performance in coursework averaged 81.7% (median=81.2 and Std. Dev.=8.8) and for exam 73.4% (median=76.1 and Std. Dev.=15.2). On balance, there were no statistically significant differences across gender; however, females tended to be slightly older than males (mean 21.8 versus 21.2), males were studying for slightly longer and spending more hours each week on the course (mean 3.5 versus 3.3, and 6.3 versus 5.8, respectively), and females performed slightly better in coursework but poorer in the final exam (mean 82.1 versus 81.5 and 72.4 versus 73.8 respectively). Of note here, however, is that there is a disparity in the number of observations for males and females, so these statistics are not used to examine statistical significance between these two groups.
3. Findings

Our aim was to examine the extent to which Kahoot! influenced classroom dynamics, students’ engagement, motivation and learning. Findings from the analysis revealed four major themes related to students’ experience in the use of Kahoot! in the classroom: (1) attention and focus, (2) interaction and engagement, (3) learning and retention of knowledge, and (4) fun and enjoyment. Three themes were prevalent in the responses of the 14 participants. Moreover, the theme of fun and enjoyment was identified in the responses of 12 of the 14 participants. We examine our outcomes for each of the four themes in Sections 3.1 to 3.4, and then consider how the other details collected for respondents’ were related to these themes in Section 3.5.

3.1. Attention and Focus

All participants (14) seem to agree that the use of Kahoot! triggered positive attention and focus in the classroom. Some suggested that interacting with Kahoot! captured and sustained their attention, as well as enabled them to take a break in the lecture, and provided a point of difference.

Attention: While the use of Kahoot! itself was an enjoyable activity, students said that Kahoot!’s motivated them to pay attention during the lecture. The deployment of Kahoot! also motivated students to closely examine lecture material in order to prepare for the Kahoot! and answer questions correctly.

“I guess it keeps you more aware in a way but you’ve got to listen throughout the lecture to know what the answer is in Kahoot! which is also a good thing. So you’re always focused if you want to do well in Kahoot!”

Having a break: A major barrier to staying focused in class was the length of the lecture as well as the time of day in which the lecture took place. Our analysis revealed that 9/14 participants highlighted the importance of having a break during lectures in order to balance and sustain a desirable level of attention during lectures. They reported that Kahoot! facilitated breaks in positive ways. Ten of the 14 respondents described staying focused in a 2 hour lecture as challenging, with some describing the experience as tedious or boring. Taking a break to engage in a fun activity allowed students to feel refreshed, providing timely relief at the halfway mark of the lecture and re-energizing students for the second hour. In addition to facilitating breaks during lecture, the use of Kahoot! also created richer variation in lecture delivery, enabling a moment of fun while continuing to engage with lecture content, only in a more light-hearted way.

A point of difference: Participants referred to Kahoot! as a unique lecture experience that is enjoyable and stimulating to learning. Compared to engagement in other lectures, students mentioned that learning with Kahoot! was a rewarding lecture experience that is captivating and desirable.

“What’s been good is that it was different... it allowed people to sort of sit back and go well this isn’t how lectures usually run. So it did capture everyone’s attention straight away.”

3.2. Interaction and Engagement

Our analysis suggest that Kahoot! gave students more opportunities to interact and engage with the lecturer, peers and lecture content by providing a fun platform on which to engage. All 14 participants reported that Kahoot! positively impacted engagement in the class, and 13 of the 14 participants said that Kahoot! increased their interaction and involvement in the lectures. Key points that emerged from the data were the importance of discussions, competition and anonymity.

Interaction and discussion: Participants reported that the use of Kahoot! fostered interactivity and engagement during lectures, through answering questions, participating in quizzes, and discussions triggered by Kahoot!. The use of Kahoot! encouraged wider participation in class as opposed to conventional classrooms where discussions are often dominated by a few extraverted students. The wider student participation in the class also fostered deeper engagement in the learning environment.

“Kahoot! gives me a platform that I can express what I think ... even though it’s silent ... I still give ideas”
Kahoot! fostered wider and active student participation, and yet provided students with the opportunity to retain their most desirable personal choice of participation. Participants reported that when engaging with Kahoot! they interacted more with peers around them and with the lecturer during and after lectures than they normally would in any other lecture. Participants pointed out that with Kahoot! in the classroom, they could decide on the level of interaction that they felt comfortable with, either participating anonymously or overtly with friends, other classmates, the lecturer or with the whole class.

“Yes it made it more interactive. I supposed I don’t talk in any other class … [I talked] with my classmates more than the teacher. I probably wouldn’t have volunteered any information to the teacher. But I definitely did have more discussions in terms of the actual content with people around me than I did in other classes.”

Competition: Nine participants discussed the competitive element of Kahoot! in relation to their interaction and engagement. Many respondents liked the competitive aspect of Kahoot!, seeing it as a motivating factor to participate, encouraging them to think critically, increasing their participating energy levels and creating a lively classroom dynamic. Competition was viewed as a strong motivator, with one respondent describing how students like to ‘perform’ and another expressing their motivation to reach the top of the scoreboard and be the best in the class. Having a desire to win encouraged many students to prepare beforehand and engage with the material. It also seems to have been an icebreaker for many students, encouraging them to interact with their peers.

“...it was almost a sense of, not just competition, I want to be the best, but also comradery, hey do you think it’s also the square, oh I hit the wrong one what did you go for?”

Despite the positive experience associated with the competitive nature of Kahoot!’s utilisation, two participants felt that the use of Kahoot! had a negative competitive effect on their learning experience. They mentioned that negative aspects of competition came into play when students focused more on the competition and having fun rather than learning. In their desire to compete, some students rushed to answer questions, not taking the time to understand the questions or the answers.

“I enjoyed it, I think towards the end we probably all got a bit distracted with names and being competitive, I think sometimes you lose sight of trying to learn new things because you are just trying to win and have fun with friends instead of learning”

Anonymity: While viewed as a negative aspect of participation in technology mediated learning environments, allowing anonymity can foster deep and enriched participation. Providing anonymous participation in a learning environment can encourage wider participation as it inculcates a sense of safety and privacy (White & Dorman, 2001). The way Kahoot! was used in the course allowed students to enter a name of choice into the system each time they participated. Students could decide if they wished to remain anonymous or identify themselves. Anonymity allowed students’ to feel safer when responding to questions. It also allowed students to focus on comparing the content of Kahoot!'s and differences of opinion, rather than comparing students’ aptitudes. This encouraged participation as students were able to take part without feeling that they were being judged for answering correctly or incorrectly. Several respondents described funny names within the Kahoot! adding positively to the element of fun and social learning in game-based environments (Squire, 2011). However, this also had the potential to shift the focus away from learning as students became distracted and no longer took the Kahoot! seriously.

“...so because it’s anonymous it never creates conflict ... so if the system is anonymous that’s good for students.”

3.3. Learning and Knowledge Retention

Nine of the 14 participants stated that Kahoot! was a useful learning tool and all 14 described Kahoot! as having a positive influence on their learning. Throughout the interviews participants made positive references to how Kahoot! supported their learning. Engaging with Kahoot! during lectures helped students not only to remember previously covered material but to understand new perspectives and increase their knowledge. Knowing that there would be a Kahoot! in class also motivated several students to prepare and review material in order to do well in the Kahoot!. In particular, students
enjoyed Kahoot!s that were relevant to the course, explored complex concepts and offered insight into applications of theory. Key benefits that participants discussed were how Kahoot!s aided revision, generated discussion and helped them to retain knowledge.

“When you get a question it does help you, you’ve got to think about the answer, you’ve got to look at lectures to prepare for it... so that’s part of revision as well”

Revision: Participants felt strongly that Kahoot! could be used for revision, with 12 participants seeing Kahoot! as a useful revision tool. In fact, three participants had used Kahoot! as a revision tool for exam preparation. Participants commonly felt the best use of the tool was to review lecture content and key topics, with Kahoot!-related course content favoured over those unrelated to the course. By repeating the content in a novel way through Kahoot!, students felt they were more likely to remember the concepts. In particular, participants mentioned Kahoot!s being useful for allowing a deeper understanding of theoretical concepts. Kahoot! also offered a brief and concise understanding of the basic concepts in the course, which was then reinforced and enriched by a class discussion that encouraged more in depth thinking.

“It helped with the revising what we’d already been taught more so than actually learning the stuff because you were already asking questions about things you’d already taught us [and] I guess that does help in the long run of actually understanding”

Discussion: Participants’ responses indicated that the discussion generated by Kahoot! was often where the most valuable learning took place. Specific benefits to post-Kahoot! discussions provided perspective, highlighted diverse opinions and allowed students a chance to evaluate their knowledge in comparison to other classmates. Kahoot! and the following discussion also gave students feedback to immediately correct their own mistakes, knowing if they got an answer right or wrong, and more importantly, why. Exploring the answers and understanding why they were right or wrong generated a deeper understanding that strongly aided participants’ engagement and retention of knowledge.

“The Kahoot! itself almost seems like a fun tool to get people back engaged and then the conversation afterwards is where the learning actually occurs. You’re not actually learning from it directly but more indirectly from the discussion afterwards”

Increasing and retaining knowledge: Six participants mentioned that Kahoot! helped them remember information during and after class. A few students also felt that Kahoot! added to their knowledge, as when new information was introduced they were more likely to remember it through a Kahoot!. In terms of knowledge retention, respondents appreciated that it was a quick and simple way to refresh their memory and continue to engage with the material. Respondents indicated that within the two hour lecture a lot of material was presented to them, making it hard to retain key concepts and facts. Kahoot!s supported students to re-grasp and retain key points from within the lecture, providing a reminder of what was covered. Participants also noted that they were more likely to remember Kahoot!s that they got wrong, as they had to consider why they got the question wrong and seek to understand the correct answer.

“It’s often good to go back because then ones you got wrong, you remember them because you are like oh I got that one wrong and it’s easier to remember them”

3.4. Fun and Enjoyment

As a game-based student response system, fun and entertainment lie at the core of Kahoot!. The data showed that respondents enjoyed the Kahoot!s. Twelve participants specifically pointed out that Kahoot! was fun. The element of enjoyment and fun underlies the positive aspects of all three aforementioned themes. However, fun and enjoyment were also alluded to as being a contributor to several negative impacts of Kahoot!.

“It was a positive interest … it wasn’t a standard boring lecture where you could sit there and read the notes later on.”

The firm preference to using Kahoot! among participants was attributed to the game features. Participants said they enjoyed the game, they liked the use of it in class, and they enjoyed the course because of the Kahoot!. Further, the aspect of fun and enjoyment seems to have helped a number of students overcome barriers to interaction that they face in a typical lecture environment. Kahoot!s as
an energetic, fun, class-wide activity (that didn’t require students to identify themselves or speak in
front of the class) served as an ice-breaker for many respondents.

“It was just a fun way of interacting and learning the stuff and seeing if you knew your stuff
with the quizzes and stuff for me that was useful”

That said, two (2) participants reported a mixed response, and one (1) of the two participants
felt the aspect of fun had a negative impact. Throughout the data it is evident that striking a balance
between fun and learning is vital to effectively using Kahoot! as a valuable tool in the classroom. It
seems as though participants reported negative impacts when the focus shifted too much in either
direction. Respondents specifically described whacky or funny names in the Kahoot!s as sometimes
distracting. They also felt that Kahoot!s involving guessing were purely for the sake of having fun and
did not contribute to their learning. Only one participant specifically mentioned that they enjoyed fun
‘off-topic’ Kahoot!s, with most participants feeling such Kahoot!s were irrelevant and an inefficient
use of class time.

“It didn’t feel directed enough ... I was kind of like why are we doing this, it just seemed like
a random fun activity... I mean it’s fun but there’s not point to it in the grand scheme of
things.”

3.5. Correlational Analysis

To supplement our qualitative results trends, we tested whether the associations between participants’
demographics, overall performance in the course and perceptions of Kahoot! were statistically
significant. Pearson’s correlations revealed that participants’ gender did not significantly correlate
with the other demographics and performance information collected – (e.g., gender, year of study),
study habits (lecture and course preparation time), or overall course grade. However, a larger sample
size is required to ensure that future quantitative analyses have adequate power to obtain significant
effects. Participant age did not correlate with demographic fac
tors or study habits; however, older students were more likely to emphasise the effects of Kahoot!’s influence on their attention and focus
during lectures, \( r = .60, p < .05 \). In the absence of an association with study habits, \( r_s = -.53, -.17, ps > .1 \), year of study positively correlated with perceptions of Kahoot!’s influence on interaction and engagement, \( r = .60, p < .05 \), which further supports the long-term value of Kahoot! and other GSRSs
in higher education. That said, this outcome somewhat contradicts Squire’s (2005) findings, as more qualified students in this work consider Kahoot! relevant and useful to their learning than their lesser qualified counterparts. Perceptions of Kahoot!’s influence on attention and focus was not significantly related to length of study, weekly lecture preparation and overall course preparation, \( r_s = -.08, -.49, .26 \) respectively, \( ps > .05 \). However, consistent with our findings above, perceptions of elevated
attention and focus during Kahoot! use positively correlated with perceptions of increase learning and
knowledge retention, \( r = .58, p < .05 \). This finding provides support for Kahoot!’s overall positive
effects on learning, notwithstanding our relatively small sample size. Finally, students’ perception of
Kahoot!’s positive influence on fun and enjoyment correlated with their level of interaction and engagement when using Kahoot! during lectures, \( r = .61, p < .05 \).

4. Summary and Implications

RQ1. What effects does Kahoot! have on classroom dynamics? We observed that Kahoot! gave
students more opportunities to engage with the lecturer, peers and lecture content by providing a fun
platform on which to engage, in a way shifting the classroom dynamics. This was particularly
different to what students experienced prior to these lectures, and in other courses. Our findings
substantiate previous research in support of the use of Kahoot! in supporting our understanding of
classroom dynamics, enhanced lecturer-student interactions, and more constructive discussions with
peers (Rosas et al., 2003; Wang, 2015). Also consistent with Wang (2015), anonymity was mentioned
as noteworthy in facilitating the less willing students. However, contrary to findings of previous
research, Kahoot! use led to excessive competition which invoked negative feelings at times.

RQ2. Does the use of Kahoot! influence students’ engagement, and how? Students felt that
Kahoot! captured their focus and interest during the course, but was also timely for allowing breaks.
This was particularly necessary for the longer lecture time that was instituted. In the same vein, the need to be attentive to perform well in Kahoot! helped students to maintain interest in the lessons during lectures. Their willingness to perform was also influenced by the level of anonymity afforded by Kahoot!, which allowed students to remain focused on comparing the content of Kahoot!'s and differences of opinion, rather than comparing other students’ aptitudes. This further emphasizes the importance of GSRSs for monitoring one’s knowledge through feedback and discussion, encoding and storing this knowledge for future use (e.g., Ke, 2009; Papastergiou, 2009). These findings also somewhat contradict the idea that students only learn through trial and error when using GSRSs (Kiili, 2005).

**RQ3. To what extent does the use of Kahoot! influence students’ motivation towards learning?** Our outcomes show that Kahoot! motivated students to be engaged, and encourage interaction in the classroom. Students were motivated to be attentive on the backdrop that they wanted to perform well in Kahoot!'s. This in turn motivated students to engage with the lecturer, peers and lecture content. Kahoot! also motivated competition in the classroom, where students were driven to see their names at the top of the leader board, and thus, were more attentive during lectures and related discussions. These effects of enhanced attention and “healthy” competition are consistent with Wang’s (2015) findings.

**RQ4. How does the use of Kahoot! enrich learning according to students’ experiences?** Student conceded that Kahoot!'s use in the course had a positive impact on the knowledge and skills they attained. Students noted that the drive to increase their attention and focus and interaction and engagement strongly supported their learning in the course. This supports previously documented positive effects of GSRS use on learning (Papastergiou, 2009). When students did not perform well in Kahoot!'s, those specific Kahoot!s were used to drive revision efforts, in view of overcoming learning deficiencies. In addition, Kahoot! offered students the opportunity to focus on specific relevant content, when a large amount of materials were delivered in lectures, which, again, is consistent with Wang’s (2015) findings. However, as student assessment approaches, Kahoot! may play more of a supporting role in the revision process as students may focus more on studying lecture content than interacting with students and the lecturer.

On balance, Kahoot!'s with the highest impact on classroom dynamics, student engagement, motivation and learning seems to be those that focussed on relevant course topics, and where there is little use of excessively distracting names and students’ behaviours. In fact, consistent with Papastergiou’s (2009) findings, students noted that Kahoot! improved classroom dynamics, engagement, motivation and learning beyond what would be expected from traditional teaching methods (e.g., normal PowerPoint slides and chalk and talk). However, we were not able to quantitatively examine such differences with the data collected; we hope to do so in future work. The themes identified support the previous studies that have found a positive effect of GSRS on, for instance, classroom dynamics (Rosas et al., 2003), learning, motivation, social interaction (e.g., Papastergiou, 2009; Wang, 2015), attention (e.g., Kiili, 2005) and willingness to prepare for class (Wang, 2015).

In terms of our contributions in this work, this study shows strong rigour interpreted through the element of credibility because we provided a systematic procedure for data coding and thematic extraction that researchers can follow in the future (Cope, 2014). The findings of this study also reflect high transferability, as our results have implications for how Kahoot! and similar GSRSs (e.g., Socrative, Quizlet and Buzz!) can be successfully implemented into university lectures in the future. The results of the present study also provide guidelines as to when and for how long Kahoot! can be a useful learning tool.

Our future research will involve a large scale deployment of Kahoot! to examine the efficacy of this tool in enhancing student learning outcomes, using quasi-experimental design as well as exploring the experiences of teachers in using Kahoot! in enhancing their teaching effectiveness. We also plan to administer a web-based survey to gather quantitative evidence to triangulate our outcomes, and particularly those around the specific aspects of GSRSs that contribute to the enrichment of learning over the use of the ‘chalkboard’ or ‘PowerPoint slides’. Furthermore, there is scope to correlate our outcomes with those provided by learning analytics tools.
Acknowledgements

We would like to thank the students for their participation in the interviews, and Dr Sander Zwanenburg and Dr Grant Dick for designing and planning lessons around Kahoot!s during INFO322. This work is funded by a University of Otago Teaching Development Grant.

References

JobStar Online: Game-Based Learning on Smartphones to Promote Youth Career Education

Toru FUJIMOTOa*, Yuki FUKUYAMAb, Satoko AZAMIC & Satoru KONNOd

aCenter for Research and Development of Higher Education, The University of Tokyo, Japan
bCollege of Arts and Science, The University of Tokyo, Japan
cIndependent Learning Game Designer, Japan
dSwitch Entertainment, Japan
*tfujimt@he.u-tokyo.ac.jp

Abstract: Career education for the youth is a crucial social need and must be upgraded and made relevant to the current generation, to help the youth develop a positive attitude toward career development and prepare for emerging opportunities. Existing job search tools are not designed to change a job seeker’s focus from current job opportunities to future possibilities. We sought to address this need while appealing to the contemporary youth’s communication styles by developing and implementing an online smartphone game called JobStar Online. The game requires participants to analyze social issues and articulate future job needs. A formative user test was conducted to evaluate the impact of playing the game on participants. Results indicated that the game offered an engaging opportunity that enhanced social interaction and facilitated learning from other players. The instructional materials need refinement to help users become familiar with the game environment, but the game created a playful context for thinking about a serious topic and supported participants in practicing idea generation.

Keywords: Career education, educational games, serious games, game-based learning, smartphone game

1. Introduction

Social and environmental changes such as globalization, political conflicts, scientific and technological advances, and climate change influence our lives and significantly reshape today’s job market. Computerization and automation technology have not only made our work more productive, but have also created new types of careers while eliminating others. For instance, careers in telemarketing, hand sewing, and watch repair are expected to disappear in the near future (Frey & Osborne, 2013). On the other hand, one study estimated that 65% of today’s grade-school students could end up doing jobs that do not yet exist (Davidson, 2011). Various new careers have emerged recently, such as data science and website design. Although it is difficult to foresee future circumstances, career education for the youth is a crucial social need and must be kept up to date to prepare the youth for jobs of the future.

In Japan, various governmental initiatives to support youth career development have been carried out in schools and elsewhere. Improving career and vocational education in schools is considered a major concern (Ministry of Education, Culture, Sports, Science and Technology of Japan, 2011). Japan’s Ministry of Economy, Trade and Industry (2013) has also taken steps to address industry sectors’ human-resource needs.

In Japan, various governmental initiatives to support youth career development have been carried out in schools and elsewhere. Improving career and vocational education in schools is considered a major concern (Ministry of Education, Culture, Sports, Science and Technology of Japan, 2011). Japan’s Ministry of Economy, Trade and Industry (2013) has also taken steps to address industry sectors’ human-resource needs.

Common approaches to this problem include holding a career education seminar with professionals as guest speakers, offering internships at local companies so that students can gain work experience, and hiring a career education coordinator to plan and implement education programs in conjunction with local businesses. These efforts offer rich sources of information that help young people think about their future professions, but they have tended to focus on matching the youth with current human-resource needs and do not necessarily consider ongoing social change and jobs of the future. They also tend to rely heavily on delivering information, through such formats as lectures and texts. Current job search techniques usually involve self-analysis activities, which tend to strengthen a job...
seeker’s views about current job opportunities rather than fostering an ability to envision future possibilities.

The present career development situation in Japan is not emotionally healthy for the youth. College students in Japan tend to experience high anxiety during their job search period. Many of them become depressed, and each year some even commit suicide (Otake, 2013). Hence, it is necessary to create more opportunities to help the youth build a positive attitude toward their career. We need to offer career development activities that are more casual and playful so that students can think about their future occupations more positively.

2. Game Development

2.1 Game-based Learning for Career Education

Game-based learning has been identified as an effective approach to making learning activities playful and engaging (Klopfer, Osterweil, & Salen, 2009). Several review studies have indicated that the number of researches on game-based learning has been significantly increasing in various fields and have provided positive outcomes in terms of different types of learning objectives such as knowledge acquisition, motivational improvement, and behavioral change (Connolly et al., 2012; Hwang & Wu, 2012; Qian & Clark, 2016). Various career-focused games exist, such as the classic Game of Life, CV (Milunski, 2013), and Career Odyssey (Franklin Learning Systems, 2008). These games can provide players with useful life lessons that they may not have experienced previously (Canary, 1968), but they tend to treat jobs as stereotypical and unchangeable. Although game-based learning is considered a potentially useful tool for career education (Miller & Knippers, 1992), there has been little consideration of how it can be best used.

There have been a few attempts to use simulation games for the purpose of career guidance (Fukamachi, 2006, 2010), but again, these simulations have guided players to view existing jobs as their career models, rather than to think about emerging jobs that do not currently exist. To encourage the youth to think about their future more positively, we presumed that it might be more effective to engage them in a game-based activity that guides them to think “outside the box.”

Although it has not applied a game-based approach, the Canadian Scholarship Trust (CST) Foundation has sought to develop more future-focused forms of youth career education. Its Inspired Minds Careers 2030 initiative provides a list, compiled by researchers and futurists, of jobs that may appear in the next 10 to 20 years (CST, 2014). To make young people think about future jobs, it is important to offer this type of information. One way to engage them in such reflection is to challenge them with a game-based activity.

2.2 A Career Education Card Game: JobStar—Create Your Star Job!

Based on these notions, we first designed a card game for career education, called JobStar—Create Your Star Job! to make career education more playful. The game is designed to be easily applied in a normal classroom setting and to engage students more positively and playfully in discussions of future jobs (Fujimoto, Fukuyama, & Azami, 2015).

The game package comprises 20 job cards, 21 event cards, and 16 industry cards. Each job card contains a generic job type such as “engineer,” “designer,” or “tour guide,” with a brief description of the major roles associated with the job. The jobs were chosen based on information in publications and on websites that are widely known by the Japanese youth (Benesse, 2014; Murakami, 2003). Among the hundreds of possible careers, we selected relatively generic, popular jobs that are simple and applicable to many industries. The appropriateness of the selection was tested by means of field tests during development. The event cards illustrate issues and incidents considered likely to occur in the near future, such as “global economy,” “national financial collapse,” and “robotization.” Finally, the industry cards describe major industries such as finance, information and communications, and healthcare. Table 1 lists the cards used for the game, and Figure 1 displays examples of card design.
Table 1: List of cards used in the game.

<table>
<thead>
<tr>
<th>Job Cards</th>
<th>Event Cards</th>
<th>Industry Cards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designer/artist</td>
<td>EdTech revolution</td>
<td>Farming, forestry, and fishing</td>
</tr>
<tr>
<td>Planner</td>
<td>Natural disaster</td>
<td>Healthcare and welfare</td>
</tr>
<tr>
<td>Researcher</td>
<td>Globalization</td>
<td>Energy development</td>
</tr>
<tr>
<td>Engineer</td>
<td>Massive tourist influx</td>
<td>Education and child care</td>
</tr>
<tr>
<td>Analyst/consultant</td>
<td>Spread of robots</td>
<td>Transportation and shipping</td>
</tr>
<tr>
<td>Trainer/coach</td>
<td>Shrinking and aging population</td>
<td>Telecommunications</td>
</tr>
<tr>
<td>Chef</td>
<td>Regional conflict</td>
<td>Manufacturing and production</td>
</tr>
<tr>
<td>Tour guide</td>
<td>Upward mobility of women</td>
<td>Tourism and food</td>
</tr>
<tr>
<td>Driver/pilot</td>
<td>Offices everywhere</td>
<td>Civil engineering</td>
</tr>
<tr>
<td>Doctor/nurse</td>
<td>Cyber society</td>
<td>Mass media</td>
</tr>
<tr>
<td>Athlete</td>
<td>National financial collapse</td>
<td>Finance</td>
</tr>
<tr>
<td>Scientist</td>
<td>Overreliance on fossil fuels</td>
<td>Government</td>
</tr>
<tr>
<td>Therapist</td>
<td>Major medical advances</td>
<td>Entertainment</td>
</tr>
<tr>
<td>Buyer/dealer/sales</td>
<td>Global warming</td>
<td>International and space</td>
</tr>
<tr>
<td>Writer</td>
<td>Food insecurity</td>
<td>Beauty and fashion</td>
</tr>
<tr>
<td>Programmer</td>
<td>Transportation revolution</td>
<td>Ceremonial and religion</td>
</tr>
<tr>
<td>Companion/communicator</td>
<td>Clean energy revolution</td>
<td></td>
</tr>
<tr>
<td>Teacher/instructor</td>
<td>Smart cities</td>
<td></td>
</tr>
<tr>
<td>Operator/technician</td>
<td>Major student influx</td>
<td></td>
</tr>
<tr>
<td>Agent</td>
<td>Space colonization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decline of rural areas</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Card design examples from JobStar.

The game is designed for three to five players. During the game, players must come up with a unique job name, based on the job cards dealt to them, and develop a job profile and an explanation of how the job might be needed in a certain context in the future, as described on their event card and industry card. Through the game play, players need to conduct various learning activities such as observing the social context, comprehending the features of jobs, identifying the relevance between the target job and the social context, envisioning how the job can be effective in such a social context, and presenting the job description to other players in just a few minutes of gameplay. The new jobs created by players are the results of these higher-order thinking skills. While the information described in the cards assists players to perform these skills, players can also learn from other players during the play. Even if a player might not know how to play the game completely, he or she could get an idea quickly by going through a few game rounds.
The basic theoretical background employed in the game design is Bandura’s (1997) social learning theory. Bandura conceptualized the characteristics of learning in a social context in terms of social modeling and vicarious learning. He emphasized the importance of surrounding oneself with other people when learning, as this brings another layer of meaning to the context. Peers learn from each other through interaction in various ways, such as observing a peer’s actions or collaborating with peers. In this game, players are not only required to present their idea but also have to listen to ideas from others for several times. Although a player may be unable to comprehend a game initially, he or she can immediately play comfortably by looking at other players and following their example. Our game utilizes this advantage of playing games in a social context.

The other theoretical consideration is the notion of situated cognition (Brown et al, 1989). Instead of setting learning as a goal, the game is designed for playing with the knowledge on jobs and social contexts with other players. The game embeds a shared goal and challenges to make use of knowledge. Learning occurs incidentally in the process of interaction with the game contents and other players. By engaging players in the playful activity, the game helps them to gain clearer images regarding jobs and reframe their thoughts on future career positively.

2.3 The Game Design of JobStar Online

Although the card-game version of JobStar was successfully implemented and the evaluation indicated that it could engage players, enhance social interactions, and facilitate participants’ learning from each other, it faced the common limitation of analog games that it could be played only in a face-to-face setting. Although analog games offer various benefits, we believed that an online version could reach a broader audience and enable players to interact effectively regardless of their geographic location. Therefore, we redesigned the card game into an online version, JobStar Online.

The basic procedure of the game is as follows:

1. At the beginning of the game, each player receives an event card, an industry, and five job cards, assigned at random.

2. The player looks at his or her hand and considers what type of job will be needed in that industry if the event described on the selected event card should occur. The player must come up with a new job type based on one of the job cards he or she is holding. For example, if a player has received the “Shrinking and aging population” event card and the “Transportation and shipping” industry card, he or she would have to imagine a society in which increasing productivity has become an issue for that industry due to the decreased number of young workers.

3. After the player has developed an idea for a new type of job in the selected industry, he or she types in the definition and significance of that job in the given context.

4. After all players have submitted their job ideas, each player casts a vote for the best job created. The number of votes collected by each player represents the points gained in the session. As participants continue to play the game and gain points, they achieve higher levels (from 1 to 50) and receive titles such as Novice Job Creator or Star Job Creator.

The technological and design specifications of JobStar Online are as follows:

- The game was developed as a smartphone application available for Android OS and iOS using the Unity game engine. The main game server was built on cloud storage hosted by Amazon Web Services.
- Although the online version maintains the basic mechanics of the card game, it offers two modes of play, “Random Play” (single player) and “Live Event” (multiplayer). Players can choose the play mode on the main menu page (Figure 2).
- In Random Play, two different levels of difficulty (easy and normal) are offered so that players can practice and become accustomed to the rules at their own pace before participating in the multiplayer mode.
- In Live Event, players compete against each other synchronously, posting jobs for as many of their job cards as they wish within the time limit. When the session closes, players evaluate the entries and vote for the job card that they like best.
- An “Original Card Creation” option enables players to create their own original event and industry cards and share them with others.
On the “My Note” page, players can review their play history and reflect on the cards they have created (Figure 3).

Figure 2. Screenshots of the game application. Left: the main menu page; center: event status list page; right: live event page.

Figure 3. Left: play history page; right: example of a job created by a player.

While JobStar Online has the strength to reach many players, social learning elements that players appreciate in an analog game version’s face-to-face setting is not evident. Therefore, we designed it to be playable individually at their disposal to substitute social learning elements with practice opportunity. Once players get accustomed to the game rule, they have an opportunity to learn by referring to the result of other players in multiplayer mode. My Note page is implemented to enhance
a reflective learning opportunity by looking through one’s play history, which is difficult to offer in analog game version.

3. Evaluation

To examine the usability and the effectiveness of the game application, we conducted a formative user test in February and March 2017. Since the game application was still in the developmental phase, we evaluated it by setting up an actual play situation. We then used an online questionnaire to collect subjective feedback from the participants on the game’s usability and their experience as players. The online questionnaire comprised 15 questions on a 5-point Likert scale (1 = not applicable, 5 = applicable) regarding how participants felt about the game, how their playing experience influenced their views regarding their future career opportunities, and their impressions of the application’s usability. These questions were followed by four free-response questions permitting the participants to reflect in a more open-ended manner on the game.

Twelve participants (nine undergraduate students and three adults) were recruited with online advertisements through Facebook and Twitter. We instructed them via email on how to install the application on their personal smartphone and asked them to play the game on their own during a particular week. By collecting their game logs, we could understand how the game was played.

During the testing period, participants tried all the game functions implemented and posted the job cards created. Each participant spent between one and two hours playing, and 69 job cards were posted during the test period.

4. Results and Discussion

4.1 Impression of the Game Play

Four participants completed the online questionnaire. Even though this number was not adequate to achieve statistical significance, the results indicated that the game could provide participants with a valuable learning opportunity in some ways (Figure 4). All respondents viewed their game play experience positively as an opportunity to change their views about careers (Q5) and to learn from other participants through playing the game (Q6). Although they found playing the game to be difficult (Q2), they did not feel that it was like studying at school, and they expressed a desire to play more games like this one (Q10). Three of the four respondents found the game fun to play (Q1); similarly, three of four reported that they were able to think about future professions that their image of future careers had changed, and that they had gained a different way of thinking through playing the game (Q3, Q4, Q7).

The following open-ended comments from the participants represent their thoughts on how this game offered them a positive experience:

It was great to understand the events likely to happen in the future while reading the cards. Also, it was nice to learn about various types of jobs. The remarkable thing was that instead of just letting players memorize this information, the game enabled us to understand new opportunities spontaneously during the intellectual activity of creating a new job. (Participant B-1)

It was nice to think about future jobs and verbalize them. It was a good opportunity for me since I have been thinking that I should work on this kind of activity, but I could not have worked on it by myself. (Participant B-2)

It was great to see the cards created by others. There were excellent ideas, and they made me think that I wanted to create better ones. (Participant B-3)

After I created a convincing job card based on the context offered, I found that the job was really expected to come into existence. (Participant B-4)
4.2 The New Jobs Created During Game Play

Players came up with various interesting new job ideas among the 69 job cards posted. For example, for the context of an aging society with a massive tourist influx, a participant created the job of “walk pacemaker,” a tour guide who has expertise in healthcare and supports elderly people in maintaining a desired walking or running pace appropriate for one’s health condition while engaging in tourism. Another participant created the career of “active senior coordinator,” a tour coordinator for older people who is knowledgeable regarding local attractions and can stimulate the local economy by helping tourists to explore the best sightseeing spots. Table 2 provides additional examples.

Although players may simply be coming up with creative or even humorous ideas, their ideas could still stimulate other players to think differently. It should be noted that the game content might not strike players as entirely entertaining, as it involves serious social issues and information regarding job descriptions and industries, which are not significantly different from the contents of a social studies textbook. Moreover, thinking about jobs is not an immediate concern for all the youth; on the other hand, many are deeply worried about this issue because they have become familiar with painful and discouraging stories about job seekers through the news media. However, during the game, participants engaged playfully with others in thinking about future jobs and enjoyed their activity.

5. Conclusion

This paper has described the development and evaluation of a smartphone game for career education, JobStar Online. The user test indicated that the game could offer an engaging opportunity that enhances players’ social interactions and enables them to learn from each other while playing. The survey results revealed that the game created a playful context for thinking about a rather serious topic that the youth...
tend to be reluctant to consider. The game also supported participants in practicing idea generation and in presenting their ideas to others by promoting interaction among the participants, as the card-game version had done.

Table 2: Example of new jobs created.

<table>
<thead>
<tr>
<th>Job title</th>
<th>Description</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic safety analyst</td>
<td>Analyzes the use of public transportation by the elderly and reports to government officials on ways to improve national transportation policy</td>
<td>Aging society, transportation</td>
</tr>
<tr>
<td>Disaster recovery therapist</td>
<td>Psychiatrist or psychological counselor specializing in mental illness caused by experiencing serious disasters</td>
<td>Natural disaster, healthcare and welfare</td>
</tr>
<tr>
<td>Drone controller</td>
<td>Drone operator who specializes in maneuvering through an area affected by a natural disaster</td>
<td>Natural disaster, civil engineering</td>
</tr>
<tr>
<td>Astro therapist</td>
<td>Therapist who specializes in caring for specific issues faced by workers in outer space</td>
<td>Space colonization</td>
</tr>
<tr>
<td>Matchmaking negotiator</td>
<td>Helps the youth to find marriage partners in a rural area facing serious depopulation</td>
<td>Decline of rural areas</td>
</tr>
</tbody>
</table>

While this preliminary study indicated that the game might work well as a tool for facilitating constructive reflection on jobs of the future, it has a limitation in evaluation. Since our survey data were collected from a very limited number of participants in a formative assessment phase, the results do not provide conclusive evidence regarding the game’s effectiveness for career education. Further research with a larger population would be required to understand the impact of JobStar Online on players. Future study would include a more extensive quantitative analysis on how players change their minds about their career after playing the game.

We intend to continue improving JobStar Online as a game-based career education tool that offers youth an opportunity to imagine their future path positively. After fixing the problems identified during the user test, we expect to release an updated version of the game through the Apple App Store and Google Play Store, aiming at distribution to a wide audience. Player data will be collected and analyzed continuously throughout the next phase of the study.

Acknowledgements

This study was partially funded by a research grant from the Hayao Nakayama Foundation for Science and Technology and Culture.

References


Urakami, M (1994) A research on female students' transition from school environment to work environment- The influence of self-efficacy expectations on career decision making-. Japan Society of Youth and Adolescent Psychology, 6 40-49.


Sufficiency Economy Philosophy-Based Mobile Game Application to Promoting Sustainability Understanding based on Inquiry Learning with Everyday Life Activities

Charoenchai WONGWATKITa*, Muttakeen CHE-LEAHb & Sasithorn CHOOKAEWc, Jintana WONGTAd, Chitphon YACHULAWETKUNAKORNe & Ratthakarn NA PHATTHALUNGf

aDepartment of Computer and Information Technology, Faculty of Industrial Education and Technology, King Mongkut’s University of Technology Thonburi, Thailand
bDepartment of Science, Wiengsuwan Wittayakom School, Thailand
cDepartment of Teacher Training in Mechanical Engineering, Faculty of Technical Education, King Mongkut’s University of Technology North Bangkok, Thailand
d,e,fEngineering Science Classroom, King Mongkut’s University of Technology Thonburi, Thailand

*wongwatkit.c@gmail.com

Abstract: His Majesty King Bhumibol Adulyadej’s Sufficiency Economy Philosophy (SEP) has been named as a way of life for sustainability. In Thailand, SEP has been included in the basic education core curriculum to learn about efficiency life, reasonably and meticulously. However, the teaching of SEP only focuses on memorization with a limited example, resulting in students’ poor understanding and are unable to apply to their daily lives. With the advantages of mobile devices, this study developed a mobile game application to address such learning flaws and enhance the understanding of sustainability through the game missions of everyday life activities. The mini game stories were carefully designed with the scenarios, situations, and missions in corresponding with SEP pillars of moderation, reasonableness, and protection. Students play and learn the topic by following the inquiry-based learning approach. In addition, a quasi-experiment was conducted with primary school students in Social Study course to investigate the effectiveness, perceptions and cognitive loads of the developed game application. The findings reveal that the students who learned with the proposed game could outperform than those who learn with the traditional e-learning with no interactive activities; meanwhile, it found that those who got the high achievements performed well during the learning process with the shorter time on the developed game. Furthermore, the high achievers reveal better perceptions towards the game than the students with low success and felt not difficult to experience this new learning platform. In other words, this developed mobile game is helpful for students to improve their understanding of sustainability for efficiency life and society.

Keywords: Sufficiency Economy Philosophy, sustainability, social study, inquiry learning, mobile learning, educational game

1. Introduction

In the past decades, human society has changed rapidly according to the changing world regarding economy, culture, politics, and environment. Especially, the dissipation can cause many social problems. UNESCO’s National Economic and Social Development Plan has contained the sufficiency economy as a guideline in administering the country in seeking for the sustainability in human lives since eighth edition (Piboolsravut, 2004; Pruetipibultham, 2010). Since then, many countries all over the world have taken these concerns into account and have developed many policies and strategies to promote the sustainability in various dimensions ranging from a country scale, a society scale, to a household level.

In 1974, His Majesty King Bhumibol Adulyadej initiated Sufficiency Economy Philosophy (SEP) as a way of life for sustainability. It comprises three core pillars (moderation, reasonableness, and protection). Such philosophy has widely been accepted and applied in various context worldwide.
Furthermore, SEP has been included in the Basic Education Core Curriculum B.E. 2551 (A.D. 2008) in Thailand to teach students about life’s efficiency, reasonableness, and meticulosity. However, the present teaching only focuses on memorization, which causes boredom and doesn’t motivate students to learn; meanwhile, the application in the agricultural context is most often used as an example. Therefore, students do not understand SEP and cannot apply the knowledge into their daily lives for sustainability (Jairak et al., 2015).

In the past years, smartphones have become accessible in many schools, while many useful applications can help enhance and improve the learning. The interaction on the touch screen can help motivate learning and improve the learning performance. Moreover, the game apps are widely popular for the children as they are simple to play, interesting and exciting to children (Daungcharone, 2016). The children enjoy playing games and learning along the way if the knowledge is added. In the meantime, game-based learning has become significant as a medium in teaching and learning that designs and incorporates lessons in the game which allow students to engage in learning, playing and practicing during the learning (Pareto, 2014).

To address the abovementioned issue, we developed a mobile game application to promote students’ sustainability understanding based on SEP. Every day’s life activities in different situations were used in the game missions to help students link their personal experiences with the core pillars of SEP. Moreover, game stages were designed based on the inquiry learning process of 5 Essential Features on Inquiry, while the game learning experience was considered based on game-based learning strategy. The game story design has been confirmed by experts before the actual development. While playing, students travel to different situations, talk with nearby people, and enjoy the game missions with their interactions to enhance understanding of sustainability; accordingly, a meaningful learning feedback is provided upon the learning performance.

Therefore, to investigate the effectiveness of the developed game, an experiment has been conducted with primary-school students in a Social Study course to answer following research questions:

- Do the students who learned with the developed game have a better understanding of sustainability compared to those who learned with the traditional e-learning?
- What is the learning performance during the learning process of the students who learned with the developed game?
- What are their perceptions and cognitive loads towards the developed game?

2. Related Studies

2.1 Sufficiency Economy Philosophy (SEP)

Sufficiency Economy Philosophy (SEP) was developed based on the fundamental of Thai culture by His Majesty King Bhumibol Adulyadej in 1974. It is a model that encourages Thai people to live with moderation, prudence, and social immunity (Supadhiloke, 2013). The SEP is widely valued among Thai people. It is necessary to study the principles and concepts to understand it correctly in order to be applied appropriately to the societal living for sustainability. SEP can be applied to individual daily life in every time at all levels.

SEP is described with the following three pillars. 1) Moderation: Sufficiency in a way that is not too much or too little, according to its own ability and not distort the other, 2) Reasonableness: Decide carefully and think about the consequences of the action, and 3) Protection: Readiness for preparing for possible risk management both for themselves and others (UNESCO, 2013). When applying the three pillars, the effect will be balanced and get ready to change quickly and widely in all dimensions both of social, environment and cultural assimilation (Malikhao, 2017; Sornsri, 2016; Supadhliloke, 2013). For example, Phanchan (2009) found the SEP is a principle that is flexible in practice and practicable to all areas and situations, while Coskun, Zimmerman and Erbug (2015) revealed that living under SEP can make silence and peace. The people who insist with SEP life can away from hunger and have the right to political freedom (Elinoff, 2016). In public health, the sufficient life can make people healthy as reported by Arpanantikul, Phuphaibul and Khuwatsumrit (2017).
Based on SEP benefits on promoting sustainability in life, the authors adopt the principles of three pillars as a developing framework in this study.

2.2 Inquiry-based Learning with 5E Learning Cycle

The principle of the Five Essential Features of Inquiry (5EFI) is an instructional activity based on the inquiry learning process in order to enhance students to create self-learning experiences, reasonable and have the ability to invest knowledge using scientific processes. Bybee and Landes (1988) characterized five steps of the inquiry-based learning process as follows. 1) Engagement: It is an introduction to a lesson or a topic which arises from doubt, is a motivator for students to create a set of questions, 2) Exploration: Exploring, searching, experimenting hypothesis to determine possible information, 3) Explanation: To summarize and to present the results, 4) Elaboration: To bring the knowledge linked to prior knowledge or experiences in order to explain events, and 5) Evaluation: To evaluate learning processes and to apply knowledge to other situations.

The process of 5EFI helps students learn effectively in various context. For example, Renken, Peffer, Otrel-Cass, Girault and Chiocarriello (2016) used 5E and project-based learning to encourage learning challenges and to enhance the understanding of the scientific skills and found that students can answer in-depth and expressive questions that demonstrate the understanding of the concept. Pareto (2014) used the 5E process to understand the reasoning in mathematics, which concluded that it could help achieve deeper levels of learning that transfer outside the game.

By these reasons, the 5E inquiry learning approach was applied in delivering stories and situations as game stages to provide the meaningful and efficient instruction.

3. Game Story Design to Promoting Sustainability Understanding

In order to best promote the understanding of sustainability on mobile learning environment, the game story has been designed into three stages. Each stage incorporates with two designated game missions for a certain situation in corresponding to daily-life activities. Each mission is expected to provide a concrete understanding on at least one of the SEP pillars. Figure 1 illustrates the overall game story design as follows:

- **Stage 1**: The activity in this stage is self-occurrence as to visit the grand mom’s house with the humid condition. Taking a bath, hair wash, and toothbrush, and selecting appropriate clothes are the missions to gain the practical understanding of protection to preventing dirt and unhealthy substances onto the body, and reasoning of appropriateness at particular time and condition, respectively.
- **Stage 2**: Happening in the family at the grand mom’s house, the missions of sufficient eating and selecting the right tools for gardening are responsible for following SEP pillars, i.e. sufficiency upon individual requirements, and reasoning to avoid the damage of improper selection of tools.
- **Stage 3**: This stage represents the story at a local market as part of the society. The mission of buying eggs by considering their quality at the lower price is to promote the moderation; while, the task of selecting apples by considering their freshness, cleanliness and pest-free for healthy conditions of self and others is to encourage both reasonableness and moderation.

In addition to each game mission, the avatar (representing the player) can communicate with the person in that given situation to lead in the story more meaningfully, naturally, e.g. having a conversation with dad as looking for the gardening tools before selecting them in a mission. Furthermore, the whole game story from situations, missions, stages to the integration with SEP has been cross-validated by three experienced scholars on SEP, three people who have been applying SEP into lives for years, and three experts on game mechanics and design. This confirms a meaningful learning experience of promoting sustainability in a mobile game environment.
4. Game Development

As game and story already been designed in the previous section, the overall game application has been constructed following 5E inquiry learning approach into four modules, plus one extra module for data recording. Therefore, there are five main modules in the developed game, as the overall structure presented in Figure 2. The game was developed with Personal Edition of Unity and built as a mobile application for Android devices, hereinafter called TheLittleSatis. The detailed description and learning process of each module is explained in Table 1.

While the student is experiencing the game, the additional module, called Personal record, analyzes the learning performance on each game mission not only for recommending the ongoing feedback, but also for providing the learning summary, including time spent, pre-/post-test scores, and collected stars, as shown in Figure 4D. In each game mission, the scoring mechanism is account for 1) emoticon feedback by checking if the corrected interactions/answers over 50%, smiling feedback is given; otherwise crying feedback, and 2) collected stars in which each star represents 33.33% interactions/answers’ accuracy.
As a result of game development, TheLittleSatis has been piloted with a small group of students and suggested by abovementioned nine experts for the final improvement. Hence, the developed mobile game is ready for the experiment.

Figure 3. Example Screenshots of Main Games.

Figure 4. Example Screenshots of Game Introduction, Pre-/Post-Tests and Summary.
Table 1: Detailed explanation of game modules and learning processes.

<table>
<thead>
<tr>
<th>Module</th>
<th>Description</th>
<th>Inquiry Learning Process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Module 1 (E1, E2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story Introduction</td>
<td>As a mini-game, the student learns to control the avatar via the walk-through game story introduction by pressing the button for the desired direction of the avatar, while get acquainted with the game environment, e.g. scenes, decoration, obstruction, and interaction.</td>
<td>This initial activity of the game is designed to engage students in the game, story and the significance of sustainability in SEP. Besides, the student can explore the game introduction.</td>
</tr>
<tr>
<td>Pretest game</td>
<td>As shown in Figure 4A, the student tilts the mobile device to move the avatar to collect the correct response upon the SEP-related items as a warm-up activity. In the end, the game gives encouragement popup upon the prior performance.</td>
<td></td>
</tr>
<tr>
<td><strong>Module 2 (E3)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEP e-book</td>
<td>As shown in Figure 4C, the student can swipe each page of an e-book on screen as learning the theory in easy-to-understand infographic format presentation. The voice and subtitle narration are provided for better perceptions while the effect sound makes the learning alive.</td>
<td>The student is explained the basic knowledge of SEP theory starting from the history to the everyday activity. This makes the connection between the student’s prior knowledge and the following learning activities.</td>
</tr>
<tr>
<td><strong>Module 3 (E4 – selected main games)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mission1 from Stage1</td>
<td>Figure 3A, the avatar has some conversation with the mom ahead of visiting the grand mom’s house by moving the avatar to meet the mom. The student then walks to the bathroom. Figure 3B, the student drags the soap for cleaning body, shampoo for hair wash, and toothbrush and toothpaste for teeth cleaning. Subsequently, the popup presents the mission feedback with emoticon face and visualizes the knowledge learned in SEP pillar representation; also, morals drawn from the mission.</td>
<td>This mission promotes the understanding of personal immunity for student via the everyday bathing activity, also the knowledge of body hygiene in getting rid of everywhere germs. Note: The ending popup is prompted after each mission according to the mission’s performance and relevant knowledge of SEP. This help student to reflect what he/she has understood and either to replay or not for better understanding.</td>
</tr>
<tr>
<td>Mission2 from Stage1</td>
<td>To find the proper outfit for the humid weather condition at the grand mom’s house, the student enters the dressing room. This mission requires the student to select three out of six pieces of clothes appropriately by dragging them into the suitcase, as shown in Figure 3D. Figure 3E illustrates the ending popup for this mission.</td>
<td>This mission helps the student understand the personal reasoning in fitting with given situation and condition, also the appropriateness of wearing outfit for the particular occasion.</td>
</tr>
<tr>
<td>Mission3 from Stage2</td>
<td>Having dinner at grand mom’s house, the avatar enjoys the meal. As shown in Figure 3C, the student taps food in the dish for eating as the amount of rice is reducing and the level of consumed food is increasing shown above. The student is aware of moderate eating by stopping eating when the indicating scale lands on the green.</td>
<td>The student inquires the knowledge of sufficiency by practicing on every food consumption. Also, the knowledge of the proper manner when having meals with family members is highlighted.</td>
</tr>
</tbody>
</table>
As shown in Figure 4B, the student tilts the phone to move the car to collect the accurate response upon the question item. According to the SEP understanding, the popup displays the feedback and encouragement. Finally, this game helps the student to evaluate the knowledge of SEP and sustainability after experiencing and elaborating on the game.

5. Experiment

5.1 Experimental Design

To evaluate the effectiveness of the proposed game, an experiment has been conducted in comparison with the traditional method. In this study, a quasi-experimental research design has been employed since the participants were not randomly assigned. The experiment was conducted with 140 fourth graders from four classes at a primary school in Thailand. With a simple drawing of the class’s representatives, 76 students from two classes were recruited in the experimental group (EG), while 64 students of the other two classes in control group (CG). Note that students from four classes were taught by the same teacher, and each class learned independently due to the school’s timetable; moreover, all students have had at least one-year mobile gaming experience.

Within two periods (90 minutes) of Social Studies hours, EG learned from the developed TheLittleSatis, while CG learned from the traditional e-learning. The traditional method required the students to passively learn the similar content of SEP topic in three chapters, each account for one SEP pillar, on the text-based e-book format with some pictures and diagrams; however, there were no any interactive activities presented.

The research instruments used in this study were:

- Pretest and Posttest: each has ten multiple-choice-question items (total score = 10). Pretest and posttest are designed in parallel to examine the understanding of sustainability content learned in both groups. The content validity was approved by three SEP scholars, while the reliability was accepted with Cronbach’s $\alpha$ of 0.885.
- Perception Questionnaire: to examine the students’ perception towards TheLittleSatis with 26 5-point Likert scale rating items on four dimensions, as shown in Table 4. The students rated each item from 1 (lowest) to 5 (highest). The questionnaire was adapted from Wongta et al. (2016), and Wongwatkit, Tekaw, Kanjana and Khrutthaka (2015) and tested for the acceptable reliability with Cronbach’s $\alpha$ of 0.822.
- Cognitive load questionnaire: to examine students’ mental loads and mental efforts towards TheLittleSatis with eight 7-point Likert scale items, ranging from 1 (not at all to me) to 7 (very true to me). The questionnaire was adapted from Hwang, Yang and Wang (2013) with Cronbach’s $\alpha$ of 0.831.

Both groups (EG and CG) of the students followed the same research process. Starting with the 10-minute orientation, the students were explained for the learning objectives and got acquainted with the materials. The students then took the pretests and experienced the assigned materials by following the embedded instructions and learning activities for 60 minutes. After a short break, the students spent about 20 minutes on post-tests, perception questionnaires, and cognitive load questionnaires; thereby the summary was given by the teacher. Moreover, the authors used the personal learning records from the application for further analysis in this study.

5.2 Experimental Results

5.2.1 Learning Performance

To examine the effectiveness of the proposed game, TheLittleSatis, the students’ learning performance of sustainability understanding was analyzed from pretest and posttest’s scores. Regardless of the effect of students’ different prior knowledge, ANCOVA test was considered by taking the pre-test scores as a covariate. Several assumptions were tested ahead. We found that the standardized residuals for the
Interventions were normally distributed, as assessed by Shapiro–Wilk’s test ($p > 0.05$). There was homogeneity of variances, as assessed by Levene’s test of homogeneity of variance ($F(1,138) = 0.296, p = 0.588$). Moreover, there was homogeneity of regression slopes as the interaction term was not statistically significant ($F(1,136) = 0.041, p = 0.752$). Therefore, ANCOVA test was performed.

Table 2 shows that there was a statistically difference between two groups of posttest scores ($F(1,137) = 13.452, p = 0.000, \eta^2 = 0.068$). This implies that TheLittleSatis could help the students to have a better understanding of sustainability.

Table 2: ANCOVA result on students’ learning performance.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>$F(1,137)$</th>
<th>(\eta^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>EG</td>
<td>76</td>
<td>8.641</td>
<td>1.561</td>
<td>8.674</td>
<td>0.147</td>
</tr>
<tr>
<td>CG</td>
<td>64</td>
<td>7.432</td>
<td>1.728</td>
<td>7.357</td>
<td>0.162</td>
</tr>
</tbody>
</table>

***$p < 0.001$

In addition to the students’ learning performance on TheLittleSatis as analyzed from the pretest and posttest scores, the collected stars could represent an ongoing understanding of sustainability on SEP core pillars, while the duration used in each mission could tell the confidence on practicing each mission. Median-split technique was performed on students’ posttest scores to categorize students into low ($n = 42$) and high ($n = 34$) groups of achievements.

Table 3 shows that high-achieving students could collect statistically more stars than those who had low success in all game stages, meaning that the high achievers learned well in between the game, and resulted in high learning performance. Meanwhile, high achievers spent significant lesser time in all stages than those who had low achievement, meaning that they were confident with the game story.

Table 3: Students’ collected stars and duration spent on TheLittleSatis on low- and high- achievement.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Collected Star (M ± SD)</th>
<th>Duration in seconds (M ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low ($n = 42$)</td>
<td>High ($n = 34$)</td>
</tr>
<tr>
<td>1</td>
<td>1.559 ± 2.144</td>
<td>2.275 ± 1.633</td>
</tr>
<tr>
<td>2</td>
<td>2.014 ± 1.775</td>
<td>2.832 ± 1.832</td>
</tr>
<tr>
<td>3</td>
<td>1.708 ± 1.401</td>
<td>2.365 ± 0.966</td>
</tr>
<tr>
<td>Total</td>
<td>5.281 ± 1.882</td>
<td>7.472 ± 1.480</td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.001; df = 74

5.2.2 Perceptions and Cognitive Loads

Based on the questionnaires’ responses from students in EG group, the results are presented in Table 4 on low- and high- achievements. For the perceptions towards TheLittleSatis application, the results ranged between high and very high on all dimensions, meaning that the students were satisfied with the game story in the application. However, it was found that students with high achievements rated significantly higher on content and presentation, which made them learned effectively than those who got low achievements.

Nevertheless, students with low results found the game generated mental loads significantly higher than those who had high achievement, implying that they confronted difficulty and challenges while learning, which could be the factor that affects the learning performance for them. While the high achievers put significantly less mental efforts to use the game, it means that they could use the app to learn comfortably as the game story was well organized.

Based on the experiment, TheLittleSatis could provide significant effectiveness on promoting the sustainability understanding compared to the traditional e-learning method. This result was supported by the benefits of game-based learning design that the missions could visualize the intangible
learning concept in the engaging way to learn, as in agreement with Gobel, Muller, Urban and Wiemeyer (2012) and Romero (2013). By playing on the mobile device, students could interact and perceive the learning concepts behind game story intuitively (Ziesemer, Müller, & Silveira, 2013). Moreover, the missions designed in each game could help students to practice in the engaging, meaningful environment, while the students with higher achievement could collect more game points and spent lesser time than those who got the lower achievement. Furthermore, the former better perceived the content presented in each game story than the latter with less difficulty and challenge to learn on the new platform. These results are in corresponding with several studies that the high achievers had the faster ability in adapting themselves to gain new knowledge on the new platform with less anxiety and concerns (Chen, Hwang, & Tsai, 2014). Consequently, the contribution of this study could provide a practical perspective on game story design in closing the learning gap with interactive, relevant game missions for a development of mobile game learning environment.

Table 4: Students’ perceptions and cognitive loads on TheLittleSatis on low- and high-achievement.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Rating (M ± SD) (Interpretation)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptions</td>
<td>Low (n = 42)</td>
<td>High (n = 34)</td>
</tr>
<tr>
<td>Content</td>
<td>4.212 ± 1.508 (high)</td>
<td>4.882 ± 1.491 (very high)</td>
</tr>
<tr>
<td>GUI/Multimedia</td>
<td>4.690 ± 2.364 (very high)</td>
<td>4.745 ± 1.022 (very high)</td>
</tr>
<tr>
<td>UX/Interaction</td>
<td>4.488 ± 1.233 (high)</td>
<td>4.541 ± 2.104 (very high)</td>
</tr>
<tr>
<td>Presentation</td>
<td>4.272 ± 0.877 (high)</td>
<td>4.852 ± 1.620 (very high)</td>
</tr>
<tr>
<td>Cognitive loads</td>
<td>Low (not all to me)</td>
<td>4.135 ± 1.669 (middle)</td>
</tr>
<tr>
<td>Mental loads</td>
<td>4.972 ± 1.787 (high)</td>
<td>3.057 ± 2.144 (low)</td>
</tr>
<tr>
<td>Mental efforts</td>
<td>4.295 ± 1.932 (middle)</td>
<td>3.057 ± 2.144 (low)</td>
</tr>
</tbody>
</table>

*p < 0.05; ***p < 0.01; df = 74

Finally, to make use of this game effectively, the authors would suggest following guidelines: 1) it is best to use on the bigger screen or tablets as some small objects in game might not be easy to interact with, 2) there should be a teacher to facilitate students while playing the game to provide more explanations and examples that are personalized to individual students, and 3) in some case with limited number of qualified devices, pair learning might be interesting alternative as they could discuss with the peer and exchange personal experience while learning. Additionally, the proposed game can be used in many schools with further investigations, e.g. gender difference, learning behavior, learning motivation, and its effect can be studied. With the limited time in developing this game version, several improvements can be made, e.g. personalized game missions to different groups of learners, a dashboard of learning reports for instruction adjustments, more games/missions to better engage and connect students in multi-languages. Moreover, data mining technique can be adopted for a better understanding of students’ usage behaviors, while formative assessment analysis could be conducted.

Acknowledgement

The authors would like to express special thanks to Preyaporn Raksasit and Supitcha Supasueb for their contributions on the game prototype design and development. Also, we would appreciate all the experts and participants who took part in the experiment.

References


The Digital Interactive Learning Theater in the Classroom for Drama-based Learning

Yu-Tzu LIU, Shang-Chiao LIN, Wei-Yi WU & Gwo-Dong CHEN*
Department of Computer Science and Information Engineering, National Central University, Taiwan
*chen@csie.ncu.edu.tw

Abstract: Learners may have better achievement in learning if they could experience the scenarios of the textbooks in a classroom. Many scholars advocate drama-based learning as the practice of situated learning. However, traditional drama-based learning would encounter several difficulties including the time-consuming scenario construction and the laborious preparation of costumes and stage props.

In this study, we propose the Digital Interactive Learning Theater (DILT), providing performers a mirror-like way of performing. The DILT is to integrate technology into drama-based education and to transform the content of the textbooks into drama scenes so that teachers and students may conduct drama performance in a regular classroom and achieve situated learning. Furthermore, we gradually introduce the interactive digital elements into the theater and expect to maintain students’ learning motivation in the long-term use of the DILT. In this study, we applied the DILT to formal English curriculum in an elementary school and examined students’ learning after long-term use. In addition, we adopted the idea of immersion in the research in order to further realize whether students would be absorbed in the situated scenarios when conducting drama-based learning.

The experiment result demonstrated that students preferred the DILT compared with the non-interactive Digital Learning Theater (DLT). The statistics on immersion also revealed that learners were maintaining their motivation in the DILT even in the second year of using the learning system. We may conclude that the interactive elements have motivated students in learning even in the long-term use of the DILT.

Keywords: drama-based learning, situated learning, Digital Learning Theater, Immersive, interactive elements

1. Introduction

In 1989, Brown, Collins, and Duguid proposed the idea of situated cognition: “knowledge is situated, being in part a product of the activity, context, and culture in which it is developed and used”. In other words, they believe that knowledge should be acquired through learners’ interaction with the environment, that learners should learn in situations, and that things in daily life can be teaching materials (Brown et al., 1989). The idea of situated cognition, as applied to education, is to have learners explore in authentic situated learning, discussing and further constructing meaningful knowledge and questions related to learners’ real life (Donovan, Bransford and Pellegrino, 1999).

There are several approaches practicing situated learning, and drama is always considered and recognized as one of the most influential ways (McCaslin, 1998). Drama-based learning is one learning approach applying drama to in-class learning. It can be traced to two of the educational ideas of French philosopher Jean-Jacques Rousseau: learning by doing and learning by dramatic doing (Courtney, 1989). Learners are not limited to any form of the theater; they perform what they have learned through role-playing and imitation (Richer and Swortzell, 1992). They may interact and learn through a certain issue. Such learning is learner-centered, and learners may develop their cognition to a certain level by participating in role play and adding incidents or situations to the drama performance (Dewey, 2004).

The Digital Learning Theater (DLT) combines drama-based learning and digital technology, providing teachers and students in conducting drama activities in a classroom, and merging the images of student performers in the scenario on a virtual stage (Luo, 2015). After one-year application of the DLT, however, both our interviews with teachers and classroom observation revealed that some of the students’ attention decreased after long-term use of the DLT. In order to maintain students’ learning interest and motivation in learning with the DLT, this study proposes to add interactive elements and
virtual props to the DLT to increase the sense of novelty and interaction. The Digital Interactive Learning Theater (DILT) is the DLT with interactive props, dynamic background and dynamic performer’s size. We anticipate that this improvement can enhance students’ learning motivation in the DILT and maintain their interest in learning with the DILT.

Moreover, in order to further examine whether students are absorbed in the situated dramas, this study adopts the idea of the sense of immersion to observe students’ conditions when learning with the DILT. Immersion is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences (Baños, Botella, Garcia-Palacios, Villa, Perpiña and Alcañiz, 2000). Some research indicated that immersive learning may bring students novelty, cause positive learning attitude, maintain learning motivation, and enhance interactions among learners (Ganskop, 2010). In other words, adding those novel or intriguing elements to virtual scenarios in the virtual environment may make learners feel inspired and highly-motivated (Dede, 1995). Therefore, this study proposes that the maintaining of students’ sense of immersion in the DILT may keep students passionate and inspire them in their learning with the DILT.

2. Related Work

2.1 Situated Classrooms and English Villages

Many teachers would lead situated learning by presenting authentic situations for students to learn. English learning, for instance, would require a situated classroom or English Villages. An English Village is for learners to experience the whole English environment. In addition to English Villages, many schools will build situated classrooms. However, the budget for constructing either English Villages or situated classrooms is terribly high, and the location is quite limited (Lan, 2015). The established scenarios cannot be changed easily and their content lacks flexibility. Learners may not be able to maintain their learning interest if they always practice with the same topic.

2.2 Traditional Drama base learning

Traditional drama-based learning puts much more emphasis on improvised performances but less on scenario construction. Building drama scenarios in regular classrooms is less feasible, yet it will become highly significant in order to have students involved in the drama performance (Sun, 2005). Besides, a drama is to have performers present the story on stage and face audience; students as performers will not be able to observe their own acting and adjust immediately.

To overcome the shortcomings of the learning methods mentioned above, we designed the Digital Interactive Learning Theater (DILT). It provides performers a mirror-like way of acting, gives a sense of reality as if performers were into scenario of the drama. At the moment of drama presentation, there is also the audience, and performers can see their own acting in the meantime and make an adjustment quickly. The DILT also solves the problem of building and switching scenarios; teachers and students can control presentation of situated scenarios, costumes, and stage props with a digital tablet. Moreover, to keep students’ learning interest in and to attract student’s long-term use of the DILT, we designed diverse interactive elements based on different learning topics in textbooks.

3. System Design and Implementation

In order to transform the context of textbooks into visible scenarios and to have students experience the situations, the system integrates various interactive modes and virtual props into the Digital Learning Theater. Teachers, as well as, students can effectively conduct digital drama activities and choose different interactive modes and virtual props based on the learning topics. It helps increase the sense of
novelty and interaction in digital dramas so that learners can maintain their fondness and preference for persistently using the digital dramatic learning system.

3.1 System Architecture

The layout of the DILT is demonstrated in Figure 1. The classroom is divided into a digital theater stage area and an auditorium. Students can perform in the digital theater stage area, and the auditorium area is equipped with a projector and screen in the classroom so that both performers and audience can watch the acting immediately.

The stage area is where student performers perform dramas. We set up Kinect, computers, and screens. Kinect will capture performers’ images and sent to the computer to merge the images into the virtual scenario, and performers can watch their own performance on screen and then examine and adjust their performance in time. The auditorium area is equipped with a projector and screen in the classroom. The computer screen in the stage area is projected to the screen in the auditorium. The app installed in pad is for instructors or students to control the flow of scenes in a drama performance. When students are performing, teachers can use a tablet to shift the scenes, display subtitles, play sound effects, and control the flow. In addition, the system supports video recording; students’ performance can be videotaped for demonstration and discussion afterward. A script writing tool is developed for teachers to transform the content of the textbook into scripts for DILT.

3.2 Introducing Interactive Elements into Digital Learning Theater

In the first version of digital theater, it contains basic elements such as background, foreground, sound and music effects. To support more elements required in a theater, the second version extends digital theater with dynamic background such as forest on fire or a magic mirror, digital props such as fans and fruits, and interaction of digital props such as snow balls. We hope that the digital theater can support fantasy scripts such as harry potter and promote the immersion and interactivity of the digital theater. For example, an army keep throwing snow balls to enemy to stop their attach in Figure 2. Another example is we can provide props such as tea cups, apples, masks, and costume for the actors so that the students can use them to perform without the need to prepare real props. Figure 3 depicts a scene that actors perform by using digital props.
4. Experiment

As mentioned above, we anticipate that students can always keep themselves intrigued and motivated with the learning modes in the Digital Interactive Learning Theater. Hence, our hypothesis are:

- Students maintain their sense of immersion on DILT if we put interactive pros and dynamic background scheme according to the scenario of the textbook.
- Students prefer using DILT to DLT without interactive elements.

4.1 Procedure

Three sixth-grade classes of an elementary school in Taoyuan City participated in this research. There were approximately 30 students in each class, and 92 participants in total. Students in each class were divided into six groups. The experiment was to apply the Digital Interactive Learning Theater to formal English curriculum, and all participants had experience in using the Digital Learning Theater when they were in the fifth grade.

Each class would conduct the Digital Interactive Learning Theater for five times (5 * 40 minutes). The textbooks were from Book VII and Book VIII of the English textbook for the sixth grade published by Kang Hsuan Educational Publishing Group. In addition, participants had to fill out the questionnaire on the sense of immersion at the beginning and the end of the school year. At the end of the second semester, they were required to answer online the questionnaire on preference. We also observed students’ interest in and impression of the Digital Interactive Learning Theater after conducting such a learning approach for two years via interviews with teachers and experimental results.

4.2 Instruments

4.2.1 Questionnaire on the Sense of Immersion

In order to judge whether students feel the immersion and interaction in situated learning, we adopted Rosa Baños’ immersion and involvement scale to design the questionnaire (Baños, Botella, García-Palacios, Villa, Perpiña and Alcañiz, 2000). The questionnaire involved three dimensions: reality judgment, internal/external correspondence, and attention/absorption. Each question was designed in accordance with the five-level scale invented by Rensis Likert, and every participant should choose the description that is the closest to his or her answers. The scaling ranged from five points (strongly agree) to one point (strongly disagree). The result was further analyzed with the Paired-Samples T Test.

4.2.2 Questionnaire on Preference

In DILT, we design interactive elements such as dynamic background, virtual props, and dynamic actor size in the scenario according to the content of the textbook. At the sixth grade, we also provide DLT without interactive elements for some topics. At the end of school year, we ask the students how they like DILT and DLT in 1-7 level, respectively. Then, we do the within group Paired-Samples T Test.

5. Results and Discussion

5.1 Results

5.1.1 Questionnaire on the Sense of Immersion

We examined the reliability of the questionnaires collected from students with Lee Cronbach’s alpha (α) coefficient. The alpha value is above 0.7, showing the high reliability of the questionnaires. There
are 27 questions in the questionnaire, and there were total 81 (11 students were not show up at the second time) valid questionnaires collected. The questions focus on four dimensions: reality judgment, internal/external correspondence, emotional involvement, and attention/absorption. The statistics of the Paired-Samples T Test is displayed in Table 1. From the data shown in Table 1, students remain the same score for immersion. The students had the same motivation of using DILT after using DILT for a year.

Table 1: The analysis of students’ sense of immersion.

<table>
<thead>
<tr>
<th></th>
<th>DLT (N=81)</th>
<th>DILT (N=81)</th>
<th>Significant Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Reality Judgement</td>
<td>3.63</td>
<td>.735</td>
<td>3.75</td>
</tr>
<tr>
<td>Internal/External Correspondence</td>
<td>3.78</td>
<td>.710</td>
<td>3.75</td>
</tr>
<tr>
<td>Attention/Absorption</td>
<td>3.55</td>
<td>.762</td>
<td>3.59</td>
</tr>
</tbody>
</table>

5.1.2 Questionnaire on Preference

There were 90 participants filling out the questionnaire, and the result demonstrates that students’ preference for the Digital Interactive Learning Theater reaches significance (see Table 2). It reveals that compared with the non-interactive digital theater, students prefer the digital theater with interactive elements and show higher interest in drama performance.

Table 2: Preference of interactive elements.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Score the results (N=90)</th>
<th>Significant Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLT</td>
<td>Mean 4.10</td>
<td>Standard Deviation 1.593</td>
</tr>
<tr>
<td>DILT</td>
<td>Mean 5.2</td>
<td>Standard Deviation 1.755</td>
</tr>
</tbody>
</table>

(***p<0.001, ** p<0.01, and * p<0.05)

5.2 Discussion

We introduced interactive elements and virtual props to the Digital Learning Theater and turned it into the Digital Interactive Learning Theater, hoping to increase students’ interest in digital drama performances. We examined our experimental results as follows to realize whether they correspond with our above-mentioned objectives:

- Students maintain their sense of immersion on DILT if we put interactive pros and dynamic background scheme according to the scenario of the textbook.

We may observe from the statistics in Table 1 that students were maintaining their interest in learning with the Digital Interactive Learning Theater in the second year of using the system. In the questions of every dimension, including attention and motivation, the average score of the sixth grade is higher than that of the fifth grade. It demonstrates that there is no decline in students’ learning motivation. Most students agree that they focus wholeheartedly on the drama performance and prefer to learn English with such an approach.

- Students prefer using DILT to DLT without interactive elements.

Judging from the statistics displayed in Table 2, we realize that students show a positive attitude toward learning with the Digital Interactive Learning Theater. Students have experienced various things, and the digital performance has become more interesting and lively. Most students prefer the
Digital Interactive Learning Theater very much. Interactive elements, as suggested in the interview with the teachers, can increase students’ interest as well as inspire students in learning. Virtual props make easier and more convenient the preparation for stage props that are difficult to obtain in reality.

6. Conclusion

This study proposes the Digital Interactive Learning Theater for teachers and students to conduct digital drama-based learning in a classroom. In the system, we installed various interactive modes and virtual props in order to enhance students’ sense of novelty and learning motivation. We chose the same classes in an elementary school; students learned with the Digital Learning Theater when they were in the fifth grade, and they used the Digital Interactive Learning Theater when they were in the sixth grade. In these two years, we designed scenarios based on the taught textbooks and supported teachers to conduct drama performing activities.

The experiment result shows that students are much interested in the Digital Interactive Learning Theater compared with the non-interactive Digital Learning Theater. Participants mostly agree that interactive elements and virtual props have made drama performances very interesting. With the interactive elements and virtual props designed according to taught textbooks, students would maintain the sense of immersion and their emotional involvement even after the long-term use of the Digital Interactive Learning Theater. In addition, as suggested in the interview with teachers, interaction can increase learners’ interest and inspire them, which indicates that the Digital Interactive Learning Theater has a positive effect on motivating students.

References

Learning Support System for Museum exhibits using Complex Body Movements
--Enhancing Sense of Immersion in Paleontological Environment

Mikihiro TOKUOKA\textsuperscript{a*}, Hiroshi MIZOGUCHI\textsuperscript{b}, Ryohei EGUSA\textsuperscript{b,c}, Shigenori INAGAKI\textsuperscript{c}, Fusako KUSUNOKI\textsuperscript{d} & Masanori SUGIMOTO\textsuperscript{e}

\textsuperscript{a}Department of Mechanical Engineering, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba-ken, Japan
\textsuperscript{b}JSPS Research Fellow, Tokyo, Japan
\textsuperscript{c}Graduate School of Human Development and Environment, Kobe University, Hyogo, Japan
\textsuperscript{d}Department of Computing, Tama Art University, Tokyo, Japan
\textsuperscript{e}Hokkaido University, Hokkaido, Japan
*tokmiki0708@gmail.com

Abstract: This paper presents a support system to assist learning in museums. Museums are important places for children to learn about science. At present, the main learning methods in museums involve the study of various displays and their explanations. The opportunities for user to observe or experience the environment about which they are learning are limited. It is difficult for children to learn and comprehend paleontological environments. Therefore, in this study, we work towards the development of a body experience-based learning support system that is applicable to museums. In an initial step toward realizing an immersive learning support system in museums, Yoshida et al. proposed a prototype system. However, users carry out all the learning with that same body movement. On the other hand, in our proposed system, users can learn about paleontology using various body movements and complicated body movements that are characteristic to paleontology. This enhances the sense of immersion in the paleontological environment by incorporating body movement as an observation behavior of paleontology. In addition, the user’s involvement in learning and motivation to participate are improved, and the effectiveness of the individuals’ learning process is improved because they can comprehend the various body movements of extinct animals. In this study, proposed immersive learning support system is applied to the fossil exhibition at a museum. In addition, the evaluation of user questionnaires were conducted with the objective of clarifying whether the proposed system can provide learners with a realistic paleontological observation experience.

Keywords: Kinect v2 sensor, Immersive, Learning Support System, Body Movements, Game

1. Introduction

Museums play an important part in the learning experience of children in the field of science and history. (Falk, 2012). However, because the primary learning methods in the museum include research on display and explanations, the opportunities to observe or experience the environment for learners are limited. In particular, it is impossible to experience paleontological environments including extinct animals and plants and their ecology (Adachi, 2013). Watching fossils and listening to explanations is insufficient for helping children to learn about such environments. In addition, children cannot actively learn with interest only with the aid of displays or verbal explanation. Overcoming these problems will qualitatively improve the scientific learning environment in museums. With regard to these problems, a system that simulates paleontological environments and transitions, which otherwise cannot be experienced in reality, will solve these problems. The system should also enhance the sense of immersion of the learners.
In order to enhance the sense of immersion, a full body interaction interface, wherein the movement of the whole body is linked to the operation of the system was previously shown to be effective (Klemmer, 2006). Yoshida et al. (2015) also developed a system that attempted to realize a full body interaction interface; however, the system could only be operated using one simple body movement such as raising of hands. Users were required to perform all actions using the same body movements. Therefore, it is difficult for individual learning contents to leave an impression on the users. Moreover, improving the involvement and motivation of users to participate is difficult because of the repetitive body movements.

We developed an immersive learning support system, called "BELONG". BELONG supports complex body movements that are characteristic to paleontology, which will aid users to learn about paleontology. The system is designed to enhance the user’s sense of immersion in the paleontological environment by incorporating body movement as an observation behavior of paleontology. In addition, the user's involvement in learning and motivation to participate are improved; thereby improving the effectiveness of an individual’s learning experience.

In this paper, we describe an immersive learning support system for the fossil exhibitions in museums. In addition, we also describe the results of our questionnaire-based experimental evaluation with the aim of clarifying whether the proposed system can provide learners with a realistic paleontological observation experience.

2. Learning Support System

2.1 BELONG

The proposed BELONG system accepts the user’s body movements as input for observational behavior. The system uses different body movements for different extinct species. The movements of the user’s whole body and the system operation are linked; thereby, enhancing the user’s sense of immersion in a paleontological environment. Learners can actively learn while having fun. A user’s sense of immersion can be improved if the system is operated in conjunction with complicated body movements, as compared to the case wherein the system is operated using simple body movements. The complicated body movement of this system is big body movement of the individual difference. For example, running and jumping are (Sakai, 2016).

The recognition of complicated body movements and various body movements should not involve attaching expensive sensors or devices on the user’s body, particularly when the application is intended for a museum environment. In this system, we utilize Microsoft’s Kinect v2 sensor, which is a range-image sensor. BELONG consists of a Kinect v2 sensor, projector, and control PC; therefore, it enables the provision of a low-cost immersive learning experience in confined spaces. The advantage of this arrangement is that it is possible to easily modify the learning content as required.

2.2 Configuration of the System

BELONG simulates paleontological environments and transitions that are otherwise impossible to experience in reality. As a first step towards the realization of this system, we are developing a system to simulate paleoecology, which particularly deals with the study of dinosaurs, based on experiences that simulate a paleontological excavation. This was based on our assumption that learners' interest would be improved when they are allowed to virtually excavate the fossils included in the current museum exhibit.

In this section, we describe the Kinect v2 sensor used to recognize the body movements of the users. Microsoft’s Kinect v2 sensor is a range-image sensor originally developed as part of indoor video-gaming system. Although it is inexpensive, the sensor can take sophisticated measurements and adjudge the user’s location. In addition, this sensor can recognize humans and the human skeleton using the library in Kinect’s software development kit for Windows. Kinect can measure the three-dimensional skeletal location of 25 points on the human body, including hands and legs, and it can identify the user’s pose or status based on these functions. This skeletal information makes it possible to recognize various body movements. Moreover, Kinect Studio and Visual Gesture Builder are used to
recognize the complicated body movements captured by the Kinect sensor (Tokuoka, 2017). Therefore, it is possible to develop a discriminator that registers the body movements we want to recognize and can accurately recognize body movements using machine learning. In this study, we adopted the movements associated with performing excavation activities at an excavation site.

Figure 1 shows the overview of this system; it comprises a Kinect v2 sensor, a control PC, and a projector. Figure 2 shows a situation while a user is experiencing the proposed system.

In the current system, leaners can learn about five dinosaurs; Tyrannosaurs, Tambaryu, Archeopteryx, Pteranodon, and Ichthyosaurus. The system initializes operation when users stand in front of the screen. First, they select the dinosaur fossil with which they want to play by using a “pushing motion”; the users move their palm in the direction of the screen. After selection, the users begin excavation of the corresponding dinosaur fossil. In this step, the Kinect v2 sensor is used to recognize movements of the user. When the excavation movements are performed at full power and excavation is successful, a video is displayed that shows the characteristics of the excavated dinosaur. The video shows the dinosaur’s habitat and size. We determined that the excavation movement will be recognized when massive and accurate movement is done at full power. If the user makes small movements, the movements are not recognized as excavation movements. However, when the users are more enthusiastic and move their body towards the screen, their sense of immersion in the paleontological environment is enhanced. In the museum, fossils are mere sightings. However, they felt virtual excavation enhanced their participation by visualizing the virtual world.

After the users successfully complete the excavation movements, separate learning for the five dinosaurs can be conducted, wherein the five contents and recognition methods are explained to the users. Figure 3 shows the recognition methods for each of the contents.

1. **Tyrannosaurs.** Users can answer quizzes about Tyrannosaurs by pushing their palm in the direction of the screen. A video commentary is provided after users answer the quiz correctly. In the event that an incorrect is given by the user, the system allows learners to make another selection. By learning about dinosaurs in a quiz-based format, the users’ willingness to participate can be enhanced.

   Next, we describe the recognition method used in the system. First, the skeletal information and coordinates of the hand are acquired by using the Kinect sensor. When the co-ordinates of the hand change with respect to the screen, the system regards it as a click. Therefore, the users can answer the questions of the quiz using body movements.

2. **Tambaryu.** Tambaryu moves in correspondence with the user’s movement. If users stop moving, Tambaryu also will stop moving. This is made possible by estimating the speed of the user’s motion. The users can learn about the size of Tambaryu and its walking speed while having fun at the same time.

   Next, we describe the recognition method. First, the skeletal information including the coordinates of the spine are acquired. The speed and direction of the user’s walking movement are estimated based on the changes in the pixel coordinates of the user’s spine on the screen between each measurement sample. Information such as the position at this time and the speed and direction of walking of a person are transmitted to the control PC, thereby making it possible for the on-screen dinosaur to follow the user’s movement.
3. **Archeopteryx**. Users can feed an Archeopteryx by using the action of gripping and opening their fists. First, they can select the bait for the Archeopteryx, which is displayed on the screen, by using a gripping action. Then, they can move this bait toward the mouth of the Archeopteryx and open their hands, and the Archeopteryx eats the bait. In this manner, they can virtually experience feeding an Archeopteryx and remember what they fed it. Next, we describe the recognition method. First, skeletal information and coordinates of the hand are acquired. When the hand coordinates are dense, the system recognizes it as an opening action. When the hand coordinates are dispersed, the system regards it as a gripping action. By combining these actions, it is possible to realize virtual gripping and movements.

4. **Pteranodon**. When users wave their hands in an animated fashion, a Pteranodon approaches the screen. A video of the Pteranodon approaching is displayed, and users feel like they are attacked by it. However, in the video, the Pteranodon does not have sufficient muscular strength to drag the users away. This will enable users to learn about the physical features of dinosaurs. This feature uses the Kinect's gesture recognition function (Tokuoka, 2017).

5. **Ichthyosaurus**. When learners move forward and backward, they can learn about the physical features of the Ichthyosaurus. When they approach the screen, a video showing the state of the stomach is displayed. When they retreat away from the screen, the video shows the full appearance of the Ichthyosaurus body. Thus, their sense of distance changes in conjunction with action of moving forward and backward. This enables the users to learn about the size and shape of the Ichthyosaurus. Next, we describe the recognition method. First, the skeletal information and coordinates of the spine are acquired. We assume that when the three-dimensional coordinates of the spine go beyond a certain threshold (determined by the distance from the screen), the dinosaur approaches the screen. When that condition is satisfied, the image that approaches the dinosaur body flows. On the other hand, suppose that the three-dimensional coordinates of the spine exceed another threshold distance away from the screen, it is assumed that it is far from the screen. When that condition is satisfied, the image is switched to the image away from the dinosaur. In this way, the positions of the users are correlated with the screen.

By enabling the users to perform various complex body movements, we were able to develop a full-body interactive interface that is more immersive.

![Figure 3. Various Body Movements](image-url)
3. Evaluation Experiment

3.1 Methods

Thirty-five participants from 5th-grade elementary school, were used as the subjects of the study. The experiments were conducted at the Tokyo University of Science. The subjects experienced the system in a sequential manner. As mentioned earlier, five types of contents pertaining extinct species of dinosaurs were prepared. The subjects began the experiments by selecting dinosaurs in the order of their interest. The number selected for each content is as follow; Tyrannosaurs: nine, Tambaryu: five, Archeopteryx: six, Pteranodon: nine, Ichthyosaurus: six. Finally, we evaluated the system using survey questionnaires. There were four question items pertaining the experience of physical movement. Each question was scored on a scale of one to four, where four corresponded to “strongly agree” and one corresponds to “completely disagree.”

3.2 Results

First, we classified the responses into positive responses of “strongly agree” and “agree,” and negative responses of “disagree” and “completely disagree.” We then analyzed the number of positive replies and neutral and negative replies using directly established calculation: $1 \times 2$ population rate inequality. Table 1 summarizes the results of the evaluation for the physical movement experience. For all types of contents, the number of positive responses for “I enjoyed the fact that I was moving my body,” “I feel attached to extinct animal I selected,” “I found it easy to use the system with my body movements,” and “I want to know about other extinct animals from the paleontological era through the system experience” exceeded the number of neutral and negative responses. In addition, a significant deviation was observed between the number of responses.

Table 1: Evaluation of the overall system experienced.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Tyrannosaurs</th>
<th>Tambaryu</th>
<th>Archeopteryx</th>
<th>Pteranodon</th>
<th>Ichthyosaurus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td>I enjoyed the fact that I was moving my body.</td>
<td>9 **</td>
<td>0</td>
<td>5 **</td>
<td>0</td>
<td>6 **</td>
</tr>
<tr>
<td>I feel attached to extinct animal I selected.</td>
<td>9 **</td>
<td>0</td>
<td>4 **</td>
<td>1</td>
<td>6 **</td>
</tr>
<tr>
<td>I found it easy to use the system with my body movements.</td>
<td>9 **</td>
<td>0</td>
<td>5 **</td>
<td>0</td>
<td>6 **</td>
</tr>
<tr>
<td>I want to know about other extinct animals from the paleontological era through the system experience.</td>
<td>9 **</td>
<td>0</td>
<td>5 **</td>
<td>0</td>
<td>6 **</td>
</tr>
</tbody>
</table>

N=35  
**P < 0.01
4. Conclusion

In this paper, we proposed a prototype of "BELONG" as the first step of realizing an immersive learning support system for the fossil exhibits in museums. In this system, users can learn about paleontology by using various complex body movements that are characteristic of paleontology. A number of children used the proposed system and answered a questionnaire survey. The results suggest that the children were enthusiastic about the system and had fun while learning. The children experienced a further desire to learn more, and their motivation to learn increased. Therefore, by using BELONG, virtual observation of the behaviors of paleontological species based on various body movements enabled learners to experience reality.

This indicates that BELONG is effective a method for providing a place of observation to learn about aspects of paleontology, such as dinosaurs, where a direct observational experience would have been impossible.

Acknowledgements

This work was supported in part by Grants-in-Aid for Scientific Research (B). The evaluation was supported by the Museum of Nature and Human Activities, Hyogo, Japan.

References


SATOYAMA: Simulating and Teaching Game Optimal for Young Children to Learn Vegetation Succession as Management of an Actual Forest

Shuya KAWAGUCHI*, Hiroshi MIZOGUCHI¹, Ryohei EGUSA¹, Yoshiaki TAKEDA², Etsuji YAMAGUCHI¹, Shigenori INAGAKI², Fusako KUSUNOKI³, Hideo FUNAOI² & Masanori SUGIMOTO⁴

¹Department of Mechanical Engineering, Tokyo University of Science, Japan
²Research Fellow of Japan Society for the Promotion of Science, Japan
³Graduate School of Human Development and Environment, Kobe University, Japan
⁴Department of Information Design, Tama Art University, Japan
⁵Faculty of Education, Soka University, Japan
⁶Graduate School of Information Science and Technology, Hokkaido University, Japan

*bdashshuya@gmail.com

Abstract: Global environmental problems have worsened in the recent years. Children need to cope with this situation by learning about global environmental issues in a realistic manner. For children to understand more about global environmental issues, it is important for them to learn from textbooks and teachers and also actually experience these learned things. However, phenomena occurring in a large time scale are difficult to experience in reality. One of them is vegetation succession. Vegetation succession occurs in a large time scale. Hence, it is difficult for children to experience changes in vegetation even if they do field work. In this study, the authors try to solve the abovementioned problem by developing a game that can support children's learning of the vegetation succession in Japan. Children can simulate the management of SATOYAMA when they play this game. SATOYAMA is a forest which has been used by humans, between nature and a village. When a player selects an action within a predetermined time, a change caused by the action occurs in SATOYAMA. The status of the managed SATOYAMA compared with the ideal SATOYAMA state is scored at the end of the game. By doing so, children can intuitively understand the state of the SATOYAMA they managed. The results obtained from 38 children, who participated in this experiment, suggested that this game would make children interested in vegetation succession and encourage learning.

Keywords: Vegetation succession, game, management, education, environmental problem, learning support

1. Introduction

Environmental problems have worsened on a global scale in the recent years. In such a situation, children must understand and experience the change of the natural environments through a realistic method. However, one of the difficulties of this trial is the fact that theoretical learning about the subject (as taught by teachers and textbooks) does not provide a real experience of the real world. As a result, the country’s natural zone, which is referred to as the SATOYAMA, has attracted attention as a place for acquiring practical knowledge of environmental problems. The important concept in understanding SATOYAMA is accompanied by a complicated mechanism about the real situation of vegetation succession. However, practical knowledge on these concepts is not provided only by merely reading textbooks and watching movies. Children cannot completely experience the real vegetation succession in the SATOYAMA-based fieldwork because it has a large time scale (e.g., several decades and several hundred years). Therefore, understanding the vegetation succession through a realistic method is difficult. A learning tool, such as a game where a child would learn about vegetation succession without
being limited by the large time scale, is important to overcome these problems. Many studies have been conducted on using a game for learning (Facer and others, 2004, Squire and Klopfer, 2007). These early studies clarified that the simulation provided by a game helped a student obtain the ability to understand the microscopic and macroscopic world. However, our educational initiative is the only one that focuses on learning an environmental problem using the vegetation succession of the SATOYAMA. The authors developed “Human SUGOROKU,” with the maintenance of the SATOYAMA as an exercise for a learner to work on an imminent environmental problem. In this game, the change of the superiority of the indicator plant by the environmental disturbance factor (e.g., felling, landslide, precipitation, etc.) is expressed in SUGOROKU form. In addition, we perform the high study of a learning document visualizing the vegetation succession of the SATOYAMA through animation (Deguchi et al., 2010; Deguchi et al., 2012). Consequently, the game raised learning will and deepened the understanding of the complicated vegetation succession. The game also improved the player’s ability of solving the problem in conjunction with the vegetation succession (Adachi, et al., 2013, Nakayama, et al., 2014, Yoshida, et al., 2015). However, the scene, where the game was set, was limited to a specific area. Large-scale ultrasonic sensors were used in this game, and learners worked on the squares arranged in the virtual world disguised as indicator plants, but this merely realized restricted immersion. Therefore, the following four observations were made:

- The competitive relationship between plants and environmental disturbance factors in various areas could not be expressed.
- The propagation of this game was difficult because of the large-scale devices used in it and its lack of portability.
- Restricted embodiment could not provide the learners with the experience of complete immersion in the virtual world.
- The game contents are not adaptable to the learners’ intelligence.

Therefore, the authors decided on the following objectives for the current study to deepen understanding and improve the problem-solving abilities of the learners in the area of vegetation succession:

- developing a wide range of contents, including main vegetation succession, in each area of Japan;
- making a portable game using mobile devices; and
- making the contents of the game adaptable to the intelligence of learners.

As the first step in this endeavor, this study presents the details of the current implementation of the game and the experiments conducted with it.

2. Current Game Implementation

2.1 Game Content

![Figure 1. SATOYAMA Management Game Screen.](image-url)
This study introduces a game, called the SATOYAMA Management Game. SATOYAMA is a forest used by humans. The SATOYAMA Management Game can simulate a player's environmental management. Figure 1 shows a game screen, where six kinds of plants arranged on the screen grow.

- small plants: *Rubus microphyllus* and *Mallotus japonicas*
- medium-sized plants: *Pinus densiflora* and *Quercus serrata*
- tall plants: *Ilex pedunclosa* and *Castanopsis* spp.

Competition exists between plants and occurs when two or more plants grow in the same place. For example, when taller and smaller plants grow in the same dense forest area, taller plants get the sunlight they need to grow, but small plants do not because the sunlight does not reach them. A deer and a pine longicorn that eat plants are used in the SATOYAMA Management Game. While adjusting the number of plants, players must prevent the plants from being devoured by the deer and the pine longicorn and manage the environment. The player must select one of the following actions within 10 s of turn to do that:

- Afforestation,
- Deforestation,
- Pesticide Application, and
- Removing Deer.

However, the player can leave it unselected. The SATOYAMA vegetation varies according to the selected action. The players manage it for 20 turns. We set it to be managed for 300 years as 15 years pass per turn. The players can experience vegetation successions in a realistic manner at the end of the game by scaling the final state of the SATOYAMA compared with the ideal environment. An ideal environment is an environment, where many medium plants grow. An ideal environment (100 points state) is used at the start of the game.

The player first clicks the start button, then the countdown begins. The vegetation change is displayed when the timer reaches 0 s. Increasing plants are surrounded by red circles, and decreasing plants turn gray. The SATOYAMA after 15 years will then be displayed. The score after repeating this circle for 20 turns is displayed at the end. Points are deductive, and SATOYAMA’s vegetation will only comprise tall plants if people do not manage it well. We set the condition at a 0 point.

The players can visually understand the state of vegetation inheritance through the SATOYAMA Management Game. Therefore, students playing this game can easily understand the interactions that occur between plants, the effect of different events in SATOYAMA, and the management of SATOYAMA. The vegetation succession is represented by the relative relationship of each plant.

Figure 2. Game Flow.
2.2 System Configuration

The system consisted of a screen, a single focal projector, a personal computer, and a mouse. A corresponding vegetation change occurs in the SATOYAMA when a player chooses the action on the screen with a mouse. These operation and control were implemented using the C# program that we developed using Visual Studio 2013. We can optionally increase and decrease the number of each plant using this program.

Table 1 shows each phenomenon and relation with the increase and decrease of the number of plants. Beginning with an ideal state, every turn changes the SATOYAMA. Medium-sized plants increase when a player chooses “Afforestation.” The number of tall plants decreases when they choose “Deforestation.” Meanwhile, the pine longicorn appears when the number of \(P.\) densiflora is above six. A deer appears when the number of \(R.\) microphyllus is above three. The outbreak is random. The influence on vegetation caused by the two creatures depends on the plant characteristic. \(P.\) densiflora markedly decreases when pine longicorn appears. In contrast, the number of small plants increases because that of medium-sized plants decreases. The player can prevent a decrease in \(P.\) densiflora when they choose “Pesticide Application.” The deer eats all kinds of plants, but likes a small plant and seedlings. Therefore, the number of small plant decreases compared to other plants when a deer appears. Meanwhile, the number of small plants increase compared to that of the other plants when a player chooses “Removing Deer.” The tall plants grow regardless of the player’s choice of action. Furthermore, the numbers of small- and medium-sized plants decrease in competition of the growing tall plants.

The gameplay is scored from the difference between the ideal and final states of the managed SATOYAMA. A penalty point method is set from 100 points and 0 point for the vegetation of the tall plants without management.

Table 1: Change of vegetation in each phenomenon.

<table>
<thead>
<tr>
<th>Plants</th>
<th>Events</th>
<th>(R.) microphyllus</th>
<th>(M.) japonicus</th>
<th>(P.) densiflora</th>
<th>(Q.) serrata</th>
<th>(I.) pedunclosa</th>
<th>Castanopsis spp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afforestation</td>
<td></td>
<td>Increase by 4</td>
<td>Increase by 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deforestation</td>
<td></td>
<td>Increase by 5</td>
<td>Increase by 4</td>
<td>Decrease by 5</td>
<td>Decrease by 6</td>
<td>Decrease by 5</td>
<td></td>
</tr>
<tr>
<td>Pine longicorn</td>
<td></td>
<td>Increase by 2</td>
<td>Increase by 3</td>
<td>Decrease by 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pesticide</td>
<td></td>
<td>Increase by 5</td>
<td>Increase by 4</td>
<td>Increase by 6</td>
<td>Decrease by 1</td>
<td>Decrease by 1</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>Deer</td>
<td>Decrease by 3</td>
<td>Decrease by 3</td>
<td>Decrease by 2</td>
<td>Decrease by 2</td>
<td>Decrease by 1</td>
<td>Decrease by 1</td>
</tr>
<tr>
<td>Removing Deer</td>
<td></td>
<td>Increase by 3</td>
<td>Increase by 3</td>
<td>Increase by 2</td>
<td>Increase by 2</td>
<td>Increase by 1</td>
<td>Increase by 1</td>
</tr>
<tr>
<td>There are tall</td>
<td>plants</td>
<td>Decrease by 3</td>
<td>Decrease by 3</td>
<td>Decrease by 2</td>
<td>Decrease by 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Experiment

3.1 Method

Thirty-eight 6th grade students from an elementary school in Kobe, Japan, participated in the experiment. The game, which simulated the management of undeveloped woodlands near populated areas, was played thrice by each participant. The game was played by one participant at a time. Each participant was asked to consider a strategy before the commencement of each game. After playing three rounds, the participants were asked to evaluate their gaming experience by rating their response to two types of multiple-choice statements.

The survey was conducted from December 19 to 22, 2016. Figure 3 shows the experiment environment.

3.2 Result

Figure 3 shows the experiment environment. The positive answer in the questionnaire result exceeded the neutral/negative answer. The number of responses showed a significant difference. This result denoted that the participants played the SATOYAMA Management Game with enthusiasm (Kawaguchi et al., 2017).

Let us now pay attention to the game score results of 36 people. Figure 4 shows a histogram showing the distribution of the first, second, and third game scores. The distribution of 36 participants in the first round was gentle, whereas the percentage of participants in the second distribution sharply increased because the game scores exceeded 50 points. The number of participants with a game score of 60 or more in the first round were 19 people, whereas the number of participants with a game score of 60 or more in the second round were 29 people. In addition, the number of participants with a game score of 90 to 99 points increased in the third time compared with the first and second times. The third-time scatter scores were larger than those of the second time but we think that it was the result of the participants performing various analysis.
4. Conclusions

This study develops and evaluates a game that simulates the management of undeveloped woodlands near populated areas. The game provides practical training by simulating a woodland management experience. The questionnaire results implied that positive opinions outnumbered neutral/negative opinions. The difference between the numbers of responses was significant. Furthermore, the score of the game shows that the number of people who take high scores increases as participants repeatedly play the game.

These results showed that the participants played the woodland management game with enthusiasm. The findings also denoted that the participants placed considerable thought into the game strategies pertaining to the simulated changes in the woodlands. We found that the game could encourage learning.

Acknowledgments

This work was supported by JSPS KAKENHI grant numbers JP26282061, JP26560129, JP15H02936, JP16H03059, and JP16H01814. The experiment was supported by Kobe Elementary School.

References

Muse: A Musically Inspired Game To Teach Arrays and Linked Lists

Vaishali SHARMA1*, Raghib MUSARRAT2*, Sridhar CHIMALAKONDA3,
& Y Raghu REDDY4

1Indian Institute of Information Technology Sri City
2Indian Institute of Information Technology Sri City
3Indian Institute of Technology Tirupati
4International Institute of Information Technology Hyderabad,
vaishali.s15@iiits.in, 2raghib.m14@iiits.in

Abstract: Data structures is an important area of knowledge in computer science that is often difficult to understand for high school students. There has been a lot of research in computer science education around the subject of programming. Hence, it becomes imperative to explore the domain of data structures as well. With this work, we have looked into an active pedagogical approach for facilitating learning of data structures in school. This approach has been implemented by means of an interactive game. The game aims to teach arrays and linked lists through activities based on the principle of compare and contrast. The innovation in the game is the use of musical notes to teach operations in arrays and linked lists. The game was evaluated with 27 students across 8th, 9th and 10th grades using a questionnaire. The results of the evaluation showed that the group of students who learnt about arrays and linked lists through the game scored an average of 81.5% of the total marks while the other group scored an average of 72% of the total marks.

Keywords: Active learning, Data Structures, Gamification

1. Introduction

In the field of computing, data structures is a fundamental course that lays a critical foundation for the rest of the curriculum and is a solid basis for industry projects. A sound understanding of data structures, hence is indispensable for any student aiming to pursue a career in this field, which is supported by many competitive programming websites such as TopCoder, CodeChef, HackerRank.

Data structures course is perceived to be a difficult subject for students, particularly the ones in high school because it is difficult for them to visualize how the data structures are structured and analyze its properties; hence they do not get the required motivation to pursue it in the way it should be (Luo, X., & Liu, B., 2011). So, there is an ardent need to improve the understanding of data structures amongst the students. Researchers have addressed this problem from several perspectives. There have been some theories and proposals which suggest that learning experience can be improved with the adoption of gamification in educational practices (Budd, T. A., 2006). This theory was applied to a competitive game for teaching data structures (Lawrence, R., 2004). Furthermore, there was another solution which addressed this concern by using games which explained the basic functionality of some of the data structures like arrays, linked lists, stacks, queues and trees (Costa, Toda, Mesquita & Brancher, 2014). Another line of research used a combination of deep parametric abstraction techniques with interactive abstraction manipulation to provide a better understanding of the details of the operations done on data structures (Ou, Vechev & Hilliges, 2015). In this paper, we experiment with the use of music to teach data structures through an innovative game, specifically focusing on using the principle of compare and contrast to teach arrays and linked lists.

We discuss an overview of related work in Section 2. We present a theoretical basis for the game in Section 3 and elaborate the design and development of the game in Sections 4 and 5 respectively. We discuss evaluation in Section 6 followed by discussions and conclusion.
2. Related Work

Gamification in education has been understood as the addition of game-like rules, interactivity, user experience to the educational setting. It has shown positive results in quite a number of domains like marketing, politics, health and fitness. Gamification can have major influences on the users in three aspects - cognition, emotion and social (Detering, Khaled, Nacke & Dixon, 2011). The 'learn by exploration' feature in games broadens the player's perception about learning, besides teaching the users the process of thinking (Detering, Khaled, Nacke & Dixon, 2011). In the game proposed in the paper, this feature of games is exploited to explain to the users the technicalities of the operations on arrays and linked lists. The user learns how to perform a certain operation on an array or a linked list right from the scratch through the method of trial and error and the element of music in it keeps the user interested throughout the process of learning. This is explained in Section 4.

Games invoke a variety of emotional experiences - from frustration to joy. Hence, students learn to see failure as an opportunity and in turn try to find out the cause of the error instead of feeling helpless about it (Lee & Hammer, 2011). This feature of gamification grips them up with a better understanding and command over the subject. We exploit this idea in the proposed game as discussed in Section 4. Lee et al. proposed the use of game-like rules system, player experiences and cultural roles to shape learners' behavior (Lee, & Hammer, 2011). It pointed out three major areas where gamification could serve as an intervention - cognition, emotional and social. The authors asserted that well-designed gamification system could help students think about schools and the process of learning differently. Denny conducted a study on the effect of a badge based achievement system on the user participation, with two groups (Denny, 2013). The activity was designed on the lines of the approach of "student pedagogy" where the students take responsibility for creating and moderating learning resources and producing peer feedback. The result was that the group having access to the badge system outperformed the other group.

Ramon Lawrence designed a method which used competitive programming with game development for an introductory data structures course (Lawrence, 2004). This method was with a class of 55 students implement the game intelligence functionality for a board game. A user survey found that 40% of the surveyed students logged onto the game site more than 20 times, and 88% logged on at least six times; thus reflecting increased motivation in the students to learn and explore more. Estevan B. Costa et al. demonstrated an interactive web learning environment, called DSLEP (Data Structure Learning Platform) to teach some of the data structures like stacks, queues, lists, arrays and trees (Costa, Toda, Mesquita & Brancher, 2014). Decker and Lawley created a game based achievement system for the undergraduate freshmen of the Computer Science course (Detering, Khaled, Nacke & Dixon, 2011). The primary intention of the system was to increase students interaction with one another, thereby promoting teamwork while trying to achieve a particular task of the game and gaining points or awards for the same. The result of this experiment was that a record-breaking 91% of the class showed an improvement in their performance.

However, to the best of our knowledge, we could not find any related work that uses music to teach data structures, which is the core idea proposed in this paper.

3. Theoretical Basis

Designing learning environments and games without theoretical basis is a futile effort (Goodyear et al., 2004). The designed game is based on active learning which can enable students to read, write, discuss, and engage in problem-solving and to maximize their potential for intellectual growth (Slavich, & Zimbardo, 2012). Muse can be considered as a practical application of this approach that teaches the operations on arrays and linked lists through music. There is a mapping between the musical notes and their corresponding tones (SouthCalMusic), (Do Re Mi Isn't Just Child's Play: How Solfege Training Can Improve Your Ear). This mapping is exploited for explaining the concepts of arrays and linked lists as shown in Table 1. The implication of this mapping is that different sequences of musical notes produce different tunes. Also, it is implicit that the various operations on arrays and linked lists of musical notes manipulate the sequence of notes differently. So, this difference in the sequences or the effect of the operations is brought out through different tunes. Hence, the similarity and the difference in the musical tunes is analogous to the difference in
arrays and linked lists.

Table 1: Mapping of the concepts in arrays and linked lists with music.

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Arrays</th>
<th>Music</th>
<th>Linked List</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessing an element</td>
<td>Element at any index of the array can be directly accessed.</td>
<td>Directly play the desired musical note.</td>
<td>The desired element can be accessed in a sequential manner.</td>
<td>All the notes up to the desired note will be played in order.</td>
</tr>
<tr>
<td>Inserting an element</td>
<td>Cannot insert an element at a given position because the size of the array is fixed.</td>
<td>The musical sequence cannot be altered here.</td>
<td>An element is inserted by first traversing to the desired position; then the appropriate links are made.</td>
<td>The new melody will have the sound of the inserted note as well.</td>
</tr>
<tr>
<td>Deleting an element</td>
<td>Cannot delete an element directly at a given position because the size of the array is fixed.</td>
<td>The musical sequence cannot be altered here as well.</td>
<td>Traverse to the position of the desired element; alter the appropriate links.</td>
<td>The new melody will not have the sound of the deleted musical note.</td>
</tr>
</tbody>
</table>

The way music is used here can be aligned with the principle of compare and contrast. It is this principle that forms the basis of the activities designed in the game to demonstrate the concepts. It has also been assessed that teaching using the principle of compare and contrast builds on the critical thinking skills of the learner naturally (Faisal, 2006). The fact that comparative thinking is a natural form of reasoning, this principle was an apt choice for the game explained in the paper.

4. Design of Game

The game teaches the various operations on arrays and linked lists along two levels, each of which discusses operations on a particular structure - the first level teaches about the operations on arrays while the second one deals with the operations on lists. The game focuses on three operations namely update, insert and delete. At each level, the player is asked to do some task associated with a certain operation and a certain structure. On completion of the task, the next level unlocks and the player moves up. The different operations of arrays and linked lists in the game are shown in Figure 1, Figure 2 and Figure 3.

The task to be performed at every level of the game is to reproduce a given musical tune in terms of its sequence of Solfege syllables - Do, Re, Mi, Fa, So, La, Ti. The player does this by rearranging the musical blocks using the allowed operations. Music serves the purpose of an error finder in the game. At each level, the player is given a reference to the tones of each of the seven Solfege syllables, with the help of which the blocks should be arranged in the correct order based on

Figure 1: Updating notes in array, linked list.

Figure 2: Inserting notes in array and linked list.

Figure 3: Deleting notes in array and linked list.
4.1 Operations on Arrays

The scene in the game opens up with a musical tone that needs to be reproduced by the player to proceed to the next level. The scene also has some musical blocks with addresses displayed, arranged in certain order. The number of musical blocks in the sequence is fixed. This implies that the memory allocated to arrays is fixed and cannot be changed later. To access any block, the players can specify the address they want to jump to and then, jump to the block at that location. Using the given three operations: insert, delete, update, the player needs to reproduce the given sequence. In the case of update operation, the players can "jump" to any element in the array which they want to change; then update it with the desired musical note from the available seven musical notes as shown in Figure 4. The ability to "jump" given to the player here relates to the fact that any element of an array can be accessed straightaway by referring to its index. The next operation is insert operation. For this operation, if the players try to insert any musical note, they are prompted with a message stating that the array is of fixed size; hence insertion of any musical note is forbidden. The last operation is delete operation. The players, when try to delete a note in an array by jumping to its position, they are prompted with a message which says that deleting a cell of memory directly in an array is forbidden because the size of the array is not allowed to be changed.

4.2 Operations on Linked Lists

The setting in this mode is same as in the "array" mode of the game. However, the players, here cannot jump to a block because the addresses of the blocks are not revealed to them. To access any block, they walk to the first block, which then forwards it to the second block and this continues until they reach the desired block. A scene in the linked list level of the game is shown in Figure 5. In this Figure, the student (player) is represented as a person holding a sword and has to rearrange the blocks using the insert operation till the sequence plays the correct music; the green links mean they are linked to another musical block while the grey ones mean that they aren't. The player has to insert the appropriate block and make links from the previous block to the current block and from the current block to the next block. In the case of update operation, they need to traverse in steps of one from the first musical block to the desired one; then replace it by selecting the appropriate note from the available seven notes. The next operation is the insert operation. For this operation, the sequence of musical blocks has some holes in it. Now, the player has to insert appropriate musical block in the holes such that the sequence sounds same as the given musical tune. Once the players reach the desired hole, they can insert the appropriate block from the available seven notes. Next, they make the link between the previous block - the current block and the current block - next block. For the delete operation, the player has to change the sequence of the musical blocks using the delete operation only so that it produces the given musical tune. Once the players reach the desired block to be deleted, they can simply can pull it away from the list and then make links between the previous block and the next block of the currently deleted block. The player finishes the task when the sequence in the linked list plays the musical tune specified in the task.

Figure 4: Updating elements in array.
5. Development

The development of the game began with the contextualization phase, which consisted of creation of activities for each of the operations on arrays and linked lists. The activity was sketched out on the principle of compare and contrast. Such an activity created an apt setting for the implementation of the concept of active learning. As for the gamification concepts, a brief explanation of it can be seen below (i) Stages: Gained by finishing activities (ii) User Profile: Used to see the statistics of utilization of the platform. The application is developed using Unity, a cross-platform game engine.

![Figure 5: Inserting missing blocks in linked list.](image)

6. Evaluation

The evaluation for the game was carried out in the school, Chinmaya Vidyalaya at Sri City, Andhra Pradesh, India. Students from grades 8th, 9th and 10th played the game and their performance was tested through a questionnaire. The questionnaire was reviewed by two teachers of the school. A total of 27 students participated in the evaluation of the game. It was ensured that the only those students participated who did not have any prior knowledge of arrays and linked lists for a fair evaluation of the hypothesis. The students were divided into two groups - one was taught arrays and linked lists through the traditional method of chalk and talk while the other was taught through the game. 14 students were in the first group while the remaining 13 were in the other group. The effectiveness of the game was analyzed through a questionnaire on arrays and linked lists. The questionnaire had 10 questions on the operations on arrays and linked lists. Some of the questions that were included in the questionnaire were – You have an array $A = \{2,3,1,4,6,7\}$. Update the third element to 8 and write the resultant array. You have a linked list $B = 2 -> 3 -> 5 -> 9 -> 10$. Insert 7 at the third position and write the new linked list. The result was that the group which learnt arrays and linked lists through the game scored an average of 81.56% of the total marks. The other group had an average of 72.2% of the total marks. A more detailed analysis of the evaluation could not be included due to space constraints. The students were also interviewed about their experience with the game. The response was a positive one and many of them were eager to play the game again. Some of them were positive about the idea of exploring other data structures through games.

7. Discussion

The problem of teaching operations on arrays and linked lists can be addressed via a variety of approaches including games. The idea of illustrating the concept of arrays and linked lists through a music-based game is substantiated by certain arguments. First, there is a clear mapping between the Solfege syllables of music, i.e. do, re, mi, fa, sol, la, ti and their corresponding tones. In addition to
this, music can be used as an error checker because the error in music can be easily noticed. This in turn helps the players to debug through the steps they take while doing the specified task in the game. Second, using music can promote user engagement and makes the learning interactive.

Our approach and game have certain limitations. The mapping of musical notes to computing concepts is based on our intuition as we did not have a musical expert in our group as of now. However, we plan to work with a musicologist to further strengthen our approach. We limited our evaluation initially with a set of 27 students but we plan to extend it with more subjects.

We see that the game can be extended to include other data structures, activities and to even other courses in computing. As a short-term future work, we plan to design activities for some of the other data structures like stack, queue, trees, heaps etc. This game could also be extended to teach some of the simple sort and search algorithms like binary search, selection sort, insertion sort etc. Based on the initial game proposed in this paper, we are also keen to explore the idea of combining music, storytelling and gamification to teach computing to school students.

8. Conclusion

The paper presented Muse as an interactive and innovative game that uses music to teach arrays and linked lists to high school students based on compare and contrast principle. The game discussed how the sequence of musical notes could be rearranged to bring out similarities and differences between arrays and linked lists. The results of evaluating the game with 27 students were promising and also show promise that the game can be extended for other data structures like stack, queue, trees, heaps and so on. The broader impact of this game was to explore the possibility of using the natural interest of music listening of people to visualize how data structures are structured and ultimately, motivate them to develop innovative applications inspired from the fundamentals.

References


Healthy Kidney: An Educational Game for Health Awareness

Sanda HAMMEDI\textsuperscript{a}, Fathi ESSALMI\textsuperscript{b} & Maiga CHANG\textsuperscript{c}

\textsuperscript{a}Higher Institute of Computer Science and Management, University of Kairouan, Tunisia  
\textsuperscript{b}Research Laboratory of Technologies of Information and Communication & Electrical Engineering (LATICE), Tunis Higher School of engineering (ENSIT), University of Tunis, Tunisia  
\textsuperscript{c}School of Computer and Information Systems, Athabasca University, Canada  
*sanda.hammedi@yahoo.fr

Abstract: The emergence of computer educational games in learning has a great impact mainly for adolescents. Educational games can be used to teach sensitive issue such as health well-being. In particular, many works point out the success of games in improving someone’s knowledge in the medical field. This paper describes an educational game “Healthy Kidney” which aims to raise people’s awareness about the prevention against chronic kidney disease. The developed educational game is validated by a classroom experiment in which 41 students from different educational levels were involved. Students were divided into two experiment groups who played two different versions of the game and one control group which underwent a traditional method of learning. Results show that students are more satisfied with the game and enjoy using it than learning by traditional methods.

Keywords: Educational game, kidney disease, awareness, prevention, adolescent

1. Introduction

Educational games are designed to simplify the learning of certain subjects. Hence, adopting educational games can be considered as a more attractive tool to combine the pleasant and the usefulness in order to facilitate the understanding of medical issues and to raise awareness toward diseases. The use of educational games helps in identifying diseases and its contributing factors and therefore deploys the prevention to follow in order to reduce the growing threat of diseases. This can also lead to improve health knowledge aiming at reducing and even avoiding the incidence of most illnesses. In particular, implementing game based learning in the medical field has been a subject of study in many works that they proved the success of games in raising the level of knowledge of the player (Goodman et al., 2006; Lieberman, 2001; Yoon & Godwin, 2007). Since most of the existing studies aim to convey information about the cure of a specific disease rather than its prevention, the lack of interest with conveying knowledge about prevention is noted. As stated by Webb and Litton (2017) that prevention is better than cure and that it is crucial to prevent one’s self against a disease before being infected.

The World Health Organization (WHO) shed the light on the importance of prevention and considers that the prevention against chronic diseases is a vital investment which helps in saving the life of 36 millions of people over the next ten years. That’s why it is crucial to establish a healthy society where essential general knowledge about health is widespread. For that reason, an educational 3D game called Healthy Kidney is developed with Unity. It is a single player game and has two main sides: learning and playing. This game is dedicated mainly for teenagers but it is helpful for people in all ages since having good health is the common interest of everyone regardless the age. The game is validated in a secondary school in order to check its effectiveness in improving one’s knowledge.

The paper is structured as follow: Section 2 talks about the review of relevant literature and Section 3 explains the implementation of the game “Healthy Kidney”. Section 4 presents the experiment and the validation of the game. Finally, Section 5 concludes the paper.
2. Literature Review

Health field is of paramount importance and games can represent an efficient option that can resort to it in this process especially for adolescences since they spend a lot of time playing. Adolescences can acquire useful information while playing. The spread of computer games among children and adolescents should be exploited by educational software to make learning more attractive and motivating (Virvou et al., 2005). Playing games is a crucial factor that contributes on the social and mental development (Amory et al., 1999). In the same way, Rieber (1996) confirms the need to play during childhood as well as the importance of using games as a powerful tool to build an interactive learning environment.

Many researchers have demonstrated the efficiency of educational games in the medical field. Bronkie the bronchiasaurus is an educational medical game which aims to teach asthma self-management (Lieberman, 2001). It consists of two asthmatic dinosaurs that the player has to control them in taking their medication and avoiding triggers. The player also has to answer questions about the respiratory system and basic asthma self-management and how to handle common social situations in asthma emergencies cases. Sickle Cell Slime-O-Rama Game is also one for cure purpose (Yoon & Godwin, 2007). It aims to improve children’s knowledge and ability in self-management for the Sickle Cell disease. This game has been tested and the results show its great utility in fostering better self-management behaviors in children.

Re-mission is designed for people who have cancer and aims to provide them cancer-related knowledge (Tale et al., 2009). Through the body of patients affected by cancer, the players can discover symptoms, destroy cancer cell, stop the spread of the disease and manage treatment’s negative effects. The game has also been evaluated and the results show its efficiency in the acquisition of cancer knowledge and the improvement of medication adherence. Dance-Dance Revolution also aims to teach about fitness and the way to burn calories for overweight people (Hoysniemi, 2006). This game promotes health by offering alternative modes of fitness and diet mode training since it instructs the player via onscreen movements that he or she has to follow and check how many calories are burnt while playing. On the other hand, Squire’s Quest! is used for users to learn the basics of prevention against numerous illnesses (Baranowski et al., 2003). It aims to increase children’s awareness about the importance of consuming fruits, vegetables and juice to prevent them from many illnesses such as cancer.

Most of the existing educational games evolve around cure rather than prevention although Webb and Litton (2017) state that prevention is better than cure and that it’s crucial to prevent one’s self against a disease before being infected. According to the World Health Organization 1, non-communicable diseases present a serious health issue that increasingly threatens the life of an appalling number of people. Among existent diseases, Chronic Kidney Disease (CKD) is chosen by the research team to implement a game to help people be aware of it and its prevention.

The proposed game should show the player what he or she has to do in order to protect himself or herself from becoming a victim of CKD. People have to be aware of the importance of their health and it’s crucial to develop a self-care behavior. Although the game can be played by everyone, its major target is the adolescents since the prevention against chronic diseases should start from the early childhood in order to avoid them in adulthood (Curtis et al., 2011). Curtis and colleagues also discussed the limited ability of many adolescents to access preventive health programs.

3. Game Implementation

The implemented game, Healthy Kidney, aims to raise the awareness of its player on how to do in order to maintain healthy kidney. It makes players acquire more knowledge in an easier and faster way while playing. Healthy Kidney is a 3D game developed with Unity. When playing the game, the player has to control his or her avatar and follows the designated path in order to reach the destination. In the route, all objects on the way can be picked up. At the same time, he or she can see the changes of the avatar’s health bar to know whether a pick-up object contributes in boosting the avatar’s physical ability or

1http://www.who.int/mediacentre/factsheets/fs355/en/
being harmful for the health. The knowledge displayed in the game are inspired from the research of Dirks and colleagues (2005) and from the French health care administration report2 (i.e., HAS).

Adding a particular feature to a game can be considered as an efficient way aiming to facilitate knowledge transmission. For instance, Bouzid and colleagues (2016) develop an educational game for deaf and people who have difficulty in hearing to enhance their vocabulary acquisition of any written language. The game is specifically based on adding a new feature (i.e., a 3D virtual signer) to the memory match game. The 3D avatar is integrated in order to facilitate learning the sign language by rendering card's content in visual-gestural modality. The 3D avatar aims to simplify the understanding of sign language by all players. Adding animated avatars to educational contexts has great advantages. Many researches have proved the success of integrating a 3D avatar in motivating the player and increasing his or her attention as well as facilitating the displaying and the understanding of knowledge in sign language (Kipp et al., 2011; Jaballah & Jemni, 2013). Many features can be added to games either for promoting knowledge understanding or for testing its usefulness in educational contexts.

The research team agreed on testing the usefulness of adding the new feature (i.e., the prompt message mechanism) to the game in order to examine its effect in enhancing knowledge acquisition. Two versions of the game are developed. The first version doesn’t prompt the health influence message about the objects on screen when the player picks them up and the second version does. The research team expects the message prompted could help players get clear idea of what they picked up and be aware of the positive or negative influence to their avatar’s health. Figure 1 shows the two versions of the game.

![Figure 1](image.png)

**Figure 1.** Screenshot of the two versions of the game which doesn't prompt messages

Both versions of the game are designated to evaluate the hypothesis that the player may not pay attention to the objects that he or she is picking up. The player actually can pick them up unintentionally. Thus, the prompted messages is reminding what objects that he or she is just picking up. As shown in Figure 1(a), the health bar increases when the avatar picks up something beneficial to the kidney and the bar decreases when the avatar picks up something harmful; for instance, there is a bottle of water that the avatar can picks up and increase its health bar. Figure 1(b) shows the second version of the game which prompts the health influence messages. As we can see, there are some fruits in front of the avatar as well as a message that appeared when it has already picked a bottle of water. It is important to note that once an object is picked up by the avatar, it will immediately be disappeared.

4. **Game Experimentation**

The research team conducted a pilot experiment at a secondary school in Kairouan city, Tunisia. Two classes with 41 students from different educational levels were recruited in the experiment. Students were randomly divided into 3 groups: two experiment groups and one control group. Each one of the two experiment groups has 13 students and the control group has 15 students. Two different learning modes were adopted in this experiment: learning through playing game and learning through traditional method. Each of the experiment groups played one of the two versions of the game to assess the

effectiveness of the game and the proposed prompt message mechanism. The control group students were learning in traditional way.

This experiment included pre-test and post-test aiming to evaluate student’s level of knowledge before and after using different methods of learning. The assessment questions cover the preventive measures against CKD (i.e., the necessity of drinking a lot of water). It aims to make player differentiate between the wrong and the right lifestyle in order to protect him/her health. These questions are formed based on the given recommendations for prevention against CKD in the research that Dirks and colleagues (2005) done and also from the French health care administration report\(^2\) (i.e., HAS). The given pre-test was composed of 20 questions to test their knowledge about CKD.

After the public health issue was introduced and a quick explanation about it was made, all students were first asked to answer some assessment questions (i.e., pre-test) to check their prior knowledge about the prevention of CKD and to know their lifestyle. The research team then gave the control group students a text containing information about CKD and explaining the preventive measures to follow; asked students in the experiment group 1 to play the first version of the game which doesn’t prompt messages; and, asked the students in the experiment group 2 to play the second version of the game which prompts messages. A post-test at the end of experiment took place to evaluate student’s knowledge learned and the results were compared to evaluate the effectiveness of different learning methods. In order to verify whether or not a hypothesis is sustained, the research team analyzes students’ performances of the pre-test and post-test as well as the difference between their pre- and post-test performances (i.e., learning gain).

The results of the post-test will make the research team know whether or not the game as well as the prompting message mechanism is useful. Our research team has the following two hypotheses:

- H1: The prompting message mechanism makes students learn better.
- H2: The games make students learn better.

Table 1 summarizes the pre- and post-test results that the three groups have. From the pre-test average performances, we can tell that students in different groups have similar understanding level for the prior knowledge about the prevention of CKD. Therefore, their learning gains could help us verify the two hypotheses this research has.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Average difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment group 1</td>
<td>0.48</td>
<td>0.92</td>
<td>0.44</td>
</tr>
<tr>
<td>Experiment group 2</td>
<td>0.48</td>
<td>0.94</td>
<td>0.46</td>
</tr>
<tr>
<td>Control group</td>
<td>0.48</td>
<td>0.55</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Results for the three groups are recapitulated. Different average learning gains between the three groups is noted. To verify whether or not this difference is significant, one-way ANOVA test is used. In Table 2, results of ANOVA test are listed.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>1.378</td>
<td>2</td>
<td>0.689</td>
<td>35.113</td>
<td>0.000*</td>
</tr>
<tr>
<td>Within groups</td>
<td>0.746</td>
<td>38</td>
<td>0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2.124</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^*\): \(p < 0.001\)

As the results listed in Table 2, the \(p\)-value is smaller than 0.001. In this case, it’s proven that there are statistically significant results. Further independent t-tests are conducted to analyze data by comparing groups with each other in order to exactly find out which groups have significant differences in terms of students’ average learning gains between each other and which ones have similar results. Students’ average learning gains between the two experiment groups shows no significant difference. As the results listed in Table 3, the \(p\)-value is larger than 0.001. In this case the results did not show any significant difference in the average learning gain between the experiment groups 1 and 2. This proves
that the proposed prompting message mechanism is not so helpful as the research team thought since students in the experiment group 2 doesn’t have significant difference in their average learning gain. Therefore, hypothesis H1 is rejected.

Table 3: t-test results of the comparison between between experiment groups 1 and 2.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-test</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment group 1</td>
<td>13</td>
<td>0.43</td>
<td>0.174</td>
<td>-0.361</td>
<td>24</td>
<td>0.722</td>
</tr>
<tr>
<td>Experiment group 2</td>
<td>13</td>
<td>0.45</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The research team also used t-test to evaluate whether or not there is a significant difference in the average learning gain between the two experiment groups and the control group. As the results listed in Table 4, p value is smaller than 0.001. The result shows that there is a significant difference in the average learning gains between experiment group 1 and the control group students. Similarly the results listed in Table 5 show that there is a significant difference (p < 0.001) in the average learning gains between experiment group 2 and the control group students; therefore, H2 is supported and the game’s effectiveness is proved.

Table 4: t-test results of the comparison between experiment group 1 and the control group.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-test</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment group 1</td>
<td>13</td>
<td>0.43</td>
<td>0.174</td>
<td>7.213</td>
<td>26</td>
<td>0.000*</td>
</tr>
<tr>
<td>Control group</td>
<td>15</td>
<td>0.06</td>
<td>0.087</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: p < 0.001

Table 5: t-test results of the comparison between experiment group 2 and the control group.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-test</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning gain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment group 2</td>
<td>13</td>
<td>0.45</td>
<td>0.15</td>
<td>8.557</td>
<td>26</td>
<td>0.000*</td>
</tr>
<tr>
<td>Control group</td>
<td>15</td>
<td>0.06</td>
<td>0.087</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: p < 0.001

Based on the experiment results, the major remarks to state are:

- The students in the two experiment groups had done a remarkable progress at the level of gaining knowledge compared to the control group students although they had all an approximate average understanding of the prior knowledge.
- Educational game is a more efficient and useful way than traditional methods while learning the knowledge of the prevention of CKD.
- The proposed prompting message mechanism is not so helpful in the game in this study.

5. Conclusion

The research team develops a Healthy Kidney game for adolescents to help them enhance their awareness toward the needed preventive measures against chronic kidney disease. This educational game is validated in a classroom with 41 students of different level of prior knowledge. A study is conducted and the collected data has been analyzed. Results show the effectiveness of adopting games as a learning tool compared to conventional methods of learning. It is also proved that the Healthy Kidney game succeeded in improving students’ knowledge about the convenient preventive measures despite of the uselessness of the proposed prompting message mechanism. Since the impact of Healthy Kidney game on gaining knowledge has been verified, the follow-up research will focus on examining student’s perception and level of satisfaction toward the game.
References


The Effects of SDE Strategy-based Computer Games on Metacognitive Awareness

Kai-Hsiang Yang*, Chin-Yen Lu & Bou-Chuan Lu
National Taipei University of Education, Taiwan
*kh.yang.sinica@gmail.com

Abstract: Some studies show that game-based learning can raise learners’ motivation to attract learners into learning activities, and students’ metacognitive awareness is an important factor of learning performance. The SDE strategy is considered to be the form of self-dialogue learning, which can help learners to reflect and monitor the self-learning states. Therefore, in this study we proposed a SDE strategy-based computer game for learning the fraction course of the fifth grade. This study adapts a quasi-experimental design, where the students were randomly assigned to two groups. The results showed that the SDE strategy was helpful to the learners’ control and reflection on the self-learning state, so that the learners could clarify the concept of fraction in the learning process, and thus achieve learning goals.

Keywords: digital based-game learning, self-explanation, metacognitive, fraction.

1. Introduction

In recent years, many researchers try to integrate information technology into education, among which the learning content and game material combined into the digital game-based learning (DGBL) towards this direction to exploration (Castellar & Looy, 2016; Wouters & Oostendorp, 2013; Hsu, 2012). The SDE strategy (Self-explanation, group-Discussion, re-Explanation) was proposed by Chang Tzyh-Lee from Team-based Learning. Through this way, the instructor can understand the relevant concepts of learners after the teaching process. And then lead students to group discussion. During the process, group members can discuss with each other or produce conflicting ideas, and their own opinions might be changed. At last, the learners once again to provide their own opinions. Thus, in this study, we proposed a SDE strategy-based computer game for learning mathematics. Moreover, an experiment has been conducted in an elementary school mathematics course to evaluate the effectiveness of the following research questions: Do students who learn with the SDE strategy-based computer game have better metacognitive awareness inclination than students without the SDE strategy?

2. Literature review

2.1 Digital Game-based Learning

Digital games are considered to be an influential learning tool because digital games can present learning resources in a diversified way and maintain a strong appeal. If the game mechanism is designed properly, it can provide an immersive and interactive environment, will be able to attract learners to thinking abstractly and help them understand the complex concept.

However, Adams & Clark’s (2014) has found that gaming practice can lead players to try some actions until the score has been raised. Because of the lack of reflection on the results, learners can’t identify their learning status, and hence usually make wrong behaviors without any enhancement in learning. For this reason, this study aims to propose a SDE strategy-based computer game to explore the effect of learners’ metacognitive awareness inclination.
2.2 SDE Strategy (Self-explanation - group-Discussion - re-Explanation)

The SDE Strategy (Self-explanation - group-Discussion - re-Explanation) is proposed by the Chang Tzyh-Lee, and is a variation of the Team-based Learning strategy. Through applying the SDE strategy, learners will enhance the self-learning and teamwork ability. Learners not only learn to be responsible for their own and the team's learning, but also to know how to work with others, and most important of all, these two abilities are required at the current era.

The research application of SDE strategy in education is applied to natural science course of the elementary school. However, the SDE strategy has not been integrated into a digital computer game; therefore, this study aims to integrating the SDE strategy into a computer game in order to raise up learner’s metacognitive awareness inclination.

3. Experimental Design

In order to evaluate the effectiveness of our proposed computer game, an experiment was conducted to compare the metacognitive awareness inclination of the students between two groups. Participants are 48 fifth grade students from a diverse public middle school at New Taipei City. The students were randomly assigned to two groups: 24 in the experimental group and 24 in the control group. The difference between two groups was that the experimental group used the SDE strategy-based computer game to help students learn math. Fig. 1 shows the computer game, called “Discover a Whale Island.”

4. Results and Discussions

4.1 Metacognitive Awareness Inclination

The results of the Paired Sample t test showed that the experimental group (t = -3.218, p = .004 < .01) had a significant difference in the metacognitive awareness inclination after the experimental learning activity. In addition, the analysis of homogeneity of within-class regression coefficient showed that the two groups had no difference with F = .029 (p = .866 > .05), implying that the homogeneity test was passed. Following that, Analysis of covariance (ANCOVA) was employed to analyze the post-questionnaire scores of the two groups by excluding the effect of the prior-questionnaire scores. Table 1 shows the ANCOVA result. The adjusted means of the experimental group and the control group are 4.048 and 3.669, respectively; moreover, the post-questionnaire scores of the two groups reached a significant level with F = 4.288 (p = .044 < .05), showing a large effect.
Table 1: ANCOVA results of the post-test for the metacognitive awareness inclination

<table>
<thead>
<tr>
<th>Variance</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Learning Attitude</td>
<td>Experimental</td>
<td>24</td>
<td>4.100</td>
<td>.565</td>
<td>4.048</td>
<td>4.288*</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>24</td>
<td>3.616</td>
<td>.856</td>
<td>3.669</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

5. Conclusions and Suggestions

Many researchers have confirmed that digital games can promote learners in the acquisition of learning concepts have a positive effect (Clark et al., 2016; Adams & Clark, 2014; Hsu & Tsai, 2011). In this study, we focus on improving the metacognitive awareness inclination of students by the proposed SDE strategy-based computer game. According to the results, we can easily see that the SDE strategy-based computer game can help learners to raise metacognitive awareness inclination after the learning activities, which means the SDE strategy can help learners to understand the concept in depth, and the process is a constructive activities of self-concept to help learners to link between new learning materials and prior knowledge, and then achieve the learning goals. On the other hand, more gaming experienced learners will be quickly grasp the rules of the game and operation ways. It is suggested that the future researchers can analyze the effect of the learners' gaming experience.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology of the Republic of China under contract number MOST 105-2511-S-152-005-MY2.

References

Scaffolding Historical Inquiry through a Collaborative Maker-based Activity

Benjamin LILLEa* & Margarida ROMEROb
aDépartement d’études sur l’enseignement et l’apprentissage, Université Laval, Canada
bLaboratoire d’Innovation et Numérique pour l’Éducation, Université Nice Sophia Antipolis, France
* benjamin.lille.1@ulaval.ca

Abstract: Teaching strategies in history education could influence the development of learners’ competencies. Lecture-based approaches on historical knowledge have limited potential in developing students’ historical thinking and critical thinking competencies while historical inquiry has been argued to be an efficient way to develop these skills. Engaging in historical thinking development through inquiry is, however, a complex and difficult process for novice students as they will potentially face didactic and cognitive limits. We wish to contribute to the didactics of history literature in proposing scaffolds for historical inquiry through a collaborative maker-based activity combined with a knowledge building (KB) tool. Considering computer-supported collaborative learning (CSCL) literature to design collaboration, we argue that core mechanics of a sandbox videogame, 3D printing and maker culture can offer scaffolds in collaborative historical inquiry as they present congruities with cognitive operations required in historical inquiry

Keywords: Historical thinking, maker-based education, game-based learning

1. Introduction

Historical thinking development through inquiry is a process with potential obstacles and students’ resistance. Historical thinking, or historical reasoning, is argued to be an analytical and critical posture where historical sources are breeding ground in establishing fact to answer historical or historiographical questions by elaborating an historical interpretation (Yelle & Déry, 2017). Historical inquiry can be cognitively demanding for novice learners. Therefore, there is a need to induce conceptual change for them to accept the relative nature of history. Limon (2001) has stressed that teaching strategies aiming at conceptual change are requiring a higher student engagement than what traditional teaching strategies can offer. Technology-enhanced learning has therefore been argued as a way that can help scaffold historical inquiry (Brush & Saye, 2008; Hicks & Doolittle, 2008; Li & Lim, 2008). Educational technologies offer affordances that can help students in adopting a more active attitude and role towards learning (Kreijns & Kirschner, 2004; Resta & Laferrière, 2007). Within educational technologies, maker technologies such as 3D printing and the maker culture has been argued as a movement that can help develop 21st century competency such as problem-solving, creativity and critical thinking (Martin, 2015). Considering the need for conceptual change, for scaffolds and for student engagement, we stress that maker-based activity with Minecraft videogame could help students in inquiring about the past. We, therefore, argue that integrating these learning activities in history teachers’ practices could help scaffold historical inquiry in order to develop students’ historical thinking. The main research question is therefore: how can a maker-based activity scaffold historical inquiry in order to develop students’ historical thinking. Within these learning activities, collaborative maker-based education, in which the learners engage in learning by making artefacts, could support the historical inquiry required to develop students’ historical thinking.

2. Digital Games Uses for Maker-Based Education

In this building activity, learners will be engaged in the collaborative creation of an historical society in order to answer an historical or historiographical question. As building an accurate historical society is complex, the building activity is considered both as a problem to solve and a way to express historical
understanding. Learners are therefore engaged in the construction of the studied society’s physical environment like a city or a medieval village while also engaged in interpreting historical sources and accounts related to the studied society. Thus, learners are investigating the complexity of historically-situated social values, economic structure as well as political and cultural contexts. Based on the LM-GM model (Arnab et al., 2015) and its extension LM-GM-LT (Patino, Proulx, & Romero, 2016) that associates gaming mechanics with learning mechanics, we reason that the chosen sandbox videogame is coherent with our pedagogical intention, as its core mechanic is constructing. Moreover, Minecraft core gaming mechanics present congruities with Cohen, Jones, Smith, & Calandra, (2016) model of maker culture mechanics based rely on building, creating, sharing and networking. These mechanics could therefore support historical thinking components such as contextualization, historical empathy, comparing and corroboration of historical accounts. We also consider the meta-gaming concept that argues that learning with digital games can happen outside the game (Bouko & Alvarez, 2016) in designing a Knowledge-Building (KB) learning tool supporting the maker-based activity. The knowledge forum (KF) is therefore considered as a meta-making space where students are learning by discussing about the challenge that is the making of a historical society through historical inquiry.

References

Technologies (pp. 211–230). IGI Global.
A Study of Design Thinking Adaptation for Maker Education Process

Peng CHEN a*, Ding LI b* & Ronghuai HUANG c*

a Department of Educational Technology, Capital Normal University, China
b College of Life Science and Chemistry, Hunan University of Technology, China
c Smart Learning Institute, Beijing Normal University, China
*dingli_ld@126.com

Abstract: Recently, the Maker Movement has attracted more attention and has been one of the fastest-growing topics, due to contemporary technical and infrastructural developments. In this paper, we described a design thinking framework in Maker course through the 3D printer. In a case study of the public course in a university in Beijing, the students’ creative process was trained and developed through learning and using design thinking. The future work is to construct an efficient framework for Maker education and to evaluate whether the design thinking is effective to cultivate and improve creative thinking.

Keywords: Maker education, design thinking, course design, 3D printing

1. Introduction

In the past few years, the Maker Movement represented a growing movement of hobbyists, thinkers, engineers, hackers, and artists committed to creatively designing and building material objects for both playful and useful ends. The rapid growth of this movement derived from advances in technology and new digital fabrication technologies that allowed the appearance of tools such as wearable computing e-textiles, robotics, 3D printing, microprocessors, and programming languages. Researchers proved that making, playing, building, and interacting with the real world were valuable ways of learning (J. Piaget, 1950; L.S. Vygotsky, 1978). There were growing interests among educators in bringing making into K-12 education to create more opportunities for students to engage in design and engineering practices. Nowadays, the Maker Movement plays an important role both inside and outside classroom, showing that it could be a part of classroom for offering a pattern of stimulation (A.R. Basawapatna, A. Repenning, C.H. Lewis, 2013; Kafai & Vasudevan, 2015; Searle & Kafai, 2015), and a part of outside activities, such as summer camps and libraries (Telhan, Kafai, Davis, Steele, & Adleberg, 2014), demonstrating that learning is feasible in any environment as long as it is well organized.

However, courses based on the Maker movement face with a series of problems such as lack of educational approaches, teaching and learning methods, and in-depth analysis of the effects. There is a need to understand how students can best utilize Maker movement strategies to achieve better learning goal, to bring up an impact on instructional approaches and lead to a more effective way of teaching and learning. In this study, we aimed to describe a design framework for the Maker course and activities by utilizing the 3D printing.

2. Framing Design thinking in Maker Course Design

Design thinking can be considered as a human-centered, creative, participative, exploratory and problem-solving process that values different perspectives of a problem (Brown, 2008; Dunne & Martin, 2006; Melles, G. and Misic, V., 2011). In our research, we adopted the Stanford d.school process of design thinking which included empathize: to understand our users; define: to define clear project objectives; ideate: to explore ideas and solutions; prototype: to build and visualize ideas and solutions; test: to review and decide. The design action plan is an iterative process, and there are deliverable outcomes and worksheets of each action phase which will give students a point of reference when
go through the process of iteration. In our research, students learn to adapt the design process for their product.

Technology enthusiasts belonged to the Makerspace movement often use communal space equipped with multiple 3D printers, laser cutters/engravers, and CNC machines. Now there is the recent availability of low-cost 3D printers, which permit “ordinary” users to create their objects at school and home. A variety of provocative research efforts in bringing 3D printing (and fabrication in general) into classrooms, have been or are being pursued (Glen Bull, Cleb Maddox, 2010).

The purpose of our study is to foster the learner’s creative thinking and design thinking experience of the product design by 3D printing in the Maker course. We focus on whether such a concept is suitable for university curriculum model, what problems will exist, how to design, and how do students experience design thinking within the product design and development, these are our investigation points.

3. Research Design

3.1 Setting and Participants

Participants (n=20; 60% females) were undergraduate students from different majors, and they were randomly separated into five groups. Students worked collaboratively to create and research maker topics under a project-based setting. The classroom layout was changeable with the movable desks and chairs, 3D printer, Arduino and wireless network work in it. They were all in one instant messaging group. All the participants in this group can share texts, graphs, audio, video, and other files.

3.2 Course Activities and Data Collection

In order to enable students to experience the maker activities effectively, we adopted the design thinking which included empathize, define, ideate, prototype and test in our course. The Course consisted of seven sessions, which includes Introduction and Maker education, Introduction of 3D printing and design thinking, SketchUp, Arduino, APP Inventor, Case study, Maker project. Participants were expected to expand at least 4 hours in a week for each session. In the final assignment, groups were required to submit a record file, including as follows: the previous group discussion records, design documents, previous prototypes’ iteration instructions, and ultimate digital artifacts and entities, personal summary, and group presentation videos. This study adopted a mixed-method approach to collect and analyze the following data: classroom observations, digital artifacts & entities, survey and individual student interviews.

4. Findings & Discussion

4.1 Class Engagement, Course Satisfaction and Reflections

The class was very lively interesting, and students were enthusiastic about it. It was delighted to see students learning fast, accepting the technology and methods very well. Some students had lots of creative ideas, and they obtained projects from the university financial Support. They said they would continue to follow the theme.

The survey aimed to analyze the students’ satisfaction with the course. All the items were about attitude on the instructional method and software application. It is found that most of the students (more than 90%) strongly liked the design thinking integrated into the course and their project process, and more than 67% to 80% strongly agreed that the learning strategies learned in the class would help them to learn better.

Student interviews offered details on the creative work in the design and development of 3D artifacts. They had a deeper understanding of the basic knowledge of 3D printing, open source hardware principles and categories, app mobile terminal development. Almost all students were able to
understand combination and complimentary about one of the design thinking core concepts (user ideas, empathy) and the Maker movement or activities (users’ ideas could make activities more targeted).

4.2 Artifacts

At the end of the semester, each group created a product with the using of 3D printing and design thinking, and these artifacts were very creative and interesting, that include Kingfisher craft, The Ecliptic Armilla, Teaching aid tools, Lamp of doom, and Three-dimensional globe.

5. Conclusion

The current research employed design thinking to develop a framework of the Maker education in 3D Printing. The main contribution of the research was multifold. First, through the combination of learning and practice, students had a good grasp of the content of 3D printing, open source hardware, and had a deep understanding of Maker spirit and idea. Second, design thinking framework was an effective teaching and learning for the 3D Printing curriculum. Third, design thinking can effectively help Makers create their projects and products, and the switch between divergent thinking and convergent thinking facilitated people get more logic and innovative ideas. In the future, we will construct an efficient framework that can integrate the Design Thinking into K-12 Maker education, and consider how to evaluate whether the design thinking in education is effective to cultivate and improve students’ creative thinking.

Acknowledgements

This study was supported by Research on Co-creation, Sharing, Promotion and Implementation of Open Educational Resources along the Belt and Road Initiative, A national general project in educational science in "the 13th Five-year Plan". 2017, BCA170073.

References


A POE Strategy-Based Gaming Approach for Mathematics Learning

Kai-Hsiang Yang*, Hsiao-Hua Chen & Bou-Chuan Lu

Department of Mathematics and Information Education,
National Taipei University of Education, Taiwan

*khyang.sinica@gmail.com

Abstract: Although many related studies have confirmed that digital game-based learning can effectively improve students' learning motivation. Some studies have reported that properly integrating learning strategies into digital games-based learning can effectively improve students' learning achievement. However, few research has been focused on the effects of learning retention. In this study, an integrated POE strategy and digital game-based learning approach has been developed for improving learning retention, and an experiment has been conducted to evaluate students' learning achievement, learning retention in a mathematics course. The experimental results show that the POE strategy-based gaming approach can significantly improve the students’ learning achievement and learning retention.

Keywords: digital game-based learning, POE strategy, learning retention.

1. Introduction

The effective and successful teaching and learning is Learners not only to acquire new knowledge but also to retain this knowledge for long periods of time after instruction (Georghiades, 2004), so the effect of learning retention is really important. Some studies have reported that digital game-based learning can promote the learning motivation of students (Hwang, Chiu, & Chen, 2015; Hsu, Tsai, & Liang, 2011). When properly integrating learning strategies with digital games-based learning can effectively improve students' learning achievement (Charsky & Ressler, 2011; Qian & Clark, 2016), but less studies has been conducted regarding the effects of learning retention. The predict-observe-explain (POE) strategy may be able to promote the conceptual understanding of lasting retention, which is integrated into the digital game-based learning, so that the learning effect of the students can be effectively improved. Therefore, this study proposes a POE strategy-based gaming approach for mathematics learning, the following research questions are investigated:

1. Do students who learn with the integrated POE strategy and digital game-based learning approach have significantly better learning achievement than students without the POE strategy?
2. Do students who learn with the integrated POE strategy and digital game-based learning approach have significantly better learning retention than students without the POE strategy?

2. Literature Review

2.1 Digital Game-based Learning

Researchers have indicated that improper lead-in of learning strategy might decrease learning motivation, increase cognitive load and cause negative learning effects (Charsky & Ressler, 2011). So it is important to integrate some appropriate learning strategies in the digital game.

2.2 The Predict-Observe-Explain (POE) Strategy

White and Gunstone (1992) have promoted the predict-observe-explain (POE) procedure as an efficient strategy for eliciting and promoting students’ science conceptions. This strategy involves students
predicting the outcome of a problem, observing the situation of the problem, and finally explaining any discrepancies between their prediction and observation. POE strategy can help students clarify their own individual ideas and effective in promoting a durable conceptual change (Küçüközer, 2013).

3. Development of an Integrated POE Strategy and Digital Game-based Learning Mode

In this study, we tried to develop a role-playing game by integrating POE strategy as part of the gaming scenarios to assist students in improving their learning performance. Fig. 1 shows the integrating POE as part of the gaming scenarios in a gaming mission.

![Figure 1. Integrating POE as part of the gaming scenarios in a gaming mission.](image)

4. Experiment Design

The subjects of this study were two classes of fifth graders of an elementary school located in New Taipei City. A quasi-experiment was designed by assigning the students in one class with 26 students to the experimental group, and those in the other class with 26 students to the control group. The experimental group learned with the POE strategy-based gaming approach, while the control group learned with the digital game without the POE strategy. The research tools in this study included a pre-test, a post-test and a delayed post-test.

5. Experimental Result

Table 1 shows the ANCOVA result of learning achievements. There was a significant difference ($F=4.14, p < .05$) between the two groups; Table 2 shows the ANCOVA result of learning retention, and there was also a significant difference ($F=12.11, p < .01$) between the two groups; that is, the students who learned with the POE strategy-based gaming approach can showed significantly better learning achievements and learning retention than those who learned with the game without POE strategy.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Adjusted mean</th>
<th>Std.error</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group</td>
<td>26</td>
<td>69.42</td>
<td>18.28</td>
<td>69.44</td>
<td>3.40</td>
<td>4.14*</td>
</tr>
<tr>
<td>control group</td>
<td>26</td>
<td>62.54</td>
<td>19.41</td>
<td>62.51</td>
<td>3.40</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
Table 2: ANCOVA result of learning retention.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Adjusted mean</th>
<th>Std.error</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group</td>
<td>26</td>
<td>70.31</td>
<td>17.44</td>
<td>70.33</td>
<td>3.71</td>
<td>12.11**</td>
</tr>
<tr>
<td>control group</td>
<td>26</td>
<td>57.46</td>
<td>14.66</td>
<td>57.44</td>
<td>3.71</td>
<td></td>
</tr>
</tbody>
</table>

**p < .01

6. Discussion and Conclusions

In this paper, a POE strategy-based gaming approach is presented. The experimental results show that the POE strategy-based gaming approach can significantly improve students’ learning achievement and learning retention. That is, the POE strategy can help students clarify their own individual ideas and effective in promoting a durable conceptual change, as indicated by several researchers (Akpinar, 2014; Küçüközer, 2011). It means that the POE strategy is properly integrated into the game. Therefore, integrated POE strategy and digital game-based learning approach is a potential and effective way to help students for better learning.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology of the Republic of China under contract number MOST 105-2511-S-152 -005 -MY2.

References


Probing in-service Teachers’ Perceptions on TPACK-G and Acceptance of GBL

Chung-Yuan HSU\textsuperscript{a}, Jyh-Chong LIANG\textsuperscript{b}, Ching Sing CHAI\textsuperscript{c} & Chin-Chung TSAI\textsuperscript{b}

\textsuperscript{a}Department of Child Care, National Pingtung University of Science and Technology, Taiwan
\textsuperscript{b}Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan
\textsuperscript{c}Learning Sciences and Technologies Academic Group, National Institute of Education, Nanyang Technological University, Singapore

*jackohio@gmail.com

Abstract: This study was to investigate how junior and senior teachers’ perceptions of the Technological Pedagogical Content Knowledge—Games (TPACK-G) differ and how their TPACK-G perceptions predict their acceptance of game-based learning. 376 in-service elementary school teachers in Taiwan were invited to answer the TPACK-G and acceptance of game-based learning questionnaires. The results show that junior elementary school teachers tended to have higher GK, GCK and GPCK than the senior. Junior elementary school teachers’ GPCK directly predicted their Attitude and Actual teaching usage, though none of the predictions were significant for the senior. Finally, the senior elementary school teachers’ GK had negative relation to their Actual teaching usage.

Keywords: TPACK, game-based learning, teacher education, games

1. Introduction

Using digital games (hereafter named games) to support teaching and learning has been receiving growing attention. Games can positively affect students’ learning motivation and outcomes, and also improve students’ critical thinking skill and problem solving ability through appropriate pedagogical design (Akcaoglu, 2014). However, most games studies investigated the impacts of games on students’ knowledge acquisition and motivation, and very few have examined technological pedagogical content knowledge toward games (TPACK-G) and acceptance of game-based learning (GBL) of teachers, the key agent in class. Furthermore, teachers’ teaching and gaming experience plays a crucial role in influencing their TPACK-G perceptions and acceptance of GBL. Prior research has identified that junior teachers tended to perceive higher self-efficacy in their TPACK-G perceptions than the senior (Hsu, Tsai, Chang & Liang, 2017). Senior teachers were less likely to adopt games in class (Li & Huang, 2016). Thus, this study paid attention to teachers’ teaching experience, and was to probe how junior and senior teachers’ perceptions of TPACK-G differ and how their TPACK-G perceptions predict their acceptance of game-based learning. Two research questions are:

1. What were junior and senior elementary school teachers’ TPACK-G perceptions?
2. How did junior and senior teachers’ TPACK-G perceptions predict their acceptance of game-based learning?

2. Methodology

2.1 Participants

The participants of this study were 376 in-service elementary school teachers (120 males and 256 females) from different geographical regions in Taiwan. The average age was 38.94 (S.D. = 6.88). Their average year of teaching experiences was 13.32 (S.D. = 7.40).
2.2 Instruments

The TPACK-G questionnaire was used in this study to probe the elementary school teachers’ confidence in TPACK-G. This instrument, 22 items in total, was developed by Hsu, Liang, Chai and Tsai (2013) according to the work of Koh, Chai and Tsai (2013) as well as Lee and Tsai (2010). The factors were game knowledge (“I can learn digital games easily”), game pedagogical knowledge (“I know how to integrate digital games into teaching”), game content knowledge (“I can tell whether the digital games represent the targeted subject matter knowledge”), and game pedagogical content knowledge (“I can select digital games to use in my classroom that enhance what I teach, how I teach and what students learn”). The reliability are .95, .96, .96, .97, .97, respectively for GK, GCK, GPK, GPCK, and overall reliability. The second questionnaire was to probe teachers’ acceptance of game-based learning. The factors were, teachers’ attitudes toward game-based learning (“I constantly play digital games”) and actual teaching usage of games (“I am using digital games in class”). Each factor had five items. The reliability is .95 for Attitude and .93 for Actual teaching usage. All items presented with a 7-point Likert scale from 1 “strongly disagree” to 7 “strongly agree”.

2.3 Data analysis

This study used confirmatory factor analysis (CFA) to examine the construct validity of the questionnaires. The t-test was applied to compare differences in junior and senior elementary school teachers’ TPACK-G perceptions and acceptance of GBL. Then, the structural relationships among the latent variables of the questionnaires were evaluated via SEM analysis for both group teachers, respectively.

3. Results

The CFA results show that a total of twenty items were retained and grouped into six factors in the model. Three to four items were remained for each factor. All measured item factor loadings are significant (p < 0.001) and higher than 0.7 that specify the relations of the questionnaire items to their posited underlying factors in one model. The reliability coefficients for these six factors ranged from 0.89 to 0.96, the average variance extracted (AVE) ranged from 0.67 to 0.87, and the composite reliability (CR) coefficients ranged from 0.89 to 0.96. The RMSEA is 0.064 and 0.052, respectively for the junior and senior elementary school teachers, suggesting a good model fit. Table 1 shows the t-test results of comparing junior and senior elementary school teachers’ TPACK-G, attitudes, and actual teaching usage. As shown, junior elementary school teachers had significantly higher “GK,” (t = 3.16, p < 0.01) “GCK” (t = 2.47, p < 0.05) and “GPCK” scores (t = 2.00, p < 0.05) than senior elementary school teachers in TPACK-G, but not in Attitude and Actual teaching usage. Given that there are significant differences between these two groups, this study further employed SEM analysis to examine the group differences in the structural models of the factors.

Table 1: comparison of junior and senior elementary school teachers’ TPACK-G, attitudes, and actual teaching usage.

<table>
<thead>
<tr>
<th></th>
<th>GK</th>
<th>GCK</th>
<th>GPK</th>
<th>GPCK</th>
<th>Attitude</th>
<th>Actual usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior</td>
<td>4.60</td>
<td>4.82</td>
<td>4.81</td>
<td>4.76</td>
<td>3.13</td>
<td>4.03</td>
</tr>
<tr>
<td>(n=193)</td>
<td>(1.41)</td>
<td>(1.09)</td>
<td>(1.11)</td>
<td>(1.20)</td>
<td>(1.60)</td>
<td>(1.23)</td>
</tr>
<tr>
<td>Senior</td>
<td>4.14</td>
<td>4.52</td>
<td>4.59</td>
<td>4.49</td>
<td>2.87</td>
<td>3.80</td>
</tr>
<tr>
<td>(n=183)</td>
<td>(1.44)</td>
<td>(1.26)</td>
<td>(1.28)</td>
<td>(1.34)</td>
<td>(1.55)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>t-test</td>
<td>3.16**</td>
<td>2.47*</td>
<td>1.73</td>
<td>2.00*</td>
<td>1.59</td>
<td>1.66</td>
</tr>
<tr>
<td>Cohen’s d</td>
<td>0.32</td>
<td>0.25</td>
<td>0.18</td>
<td>0.21</td>
<td>0.17</td>
<td>0.18</td>
</tr>
</tbody>
</table>

The goodness-of-fit (GOF) indices showed that the junior elementary school teachers’ model well explained the data. Among the fit indices, the values of Chi-square = 332.11, degree of freedom = 162, the ratio of Chi-square to degrees of freedom = 2.50, RMSEA = 0.074, GFI = 0.86, NFI = 0.93, IFI = 0.96, TLI = 0.95 and CFI = 0.96, indicated a good model fit for this model and suggested that this model was suitable for interpreting the structural relationships among the factors of TPACK-G and Acceptance of GBL in junior elementary school teachers. According to Figure 1, junior elementary school teachers’ GK was the significantly positive factor for explaining the variation in their GCK, GPK and Attitude, whereas both of junior elementary school teachers’ GCK and GPK were the
significantly positive factor for explaining the variation in their GCK. In addition, junior elementary school teachers’ GPCK was the significantly positive factor predicting both of their Attitude and Actual teaching usage. Attitude also positively related to Actual teaching usage. In addition, the fit indices, including the values of Chi-square = 275.36, degree of freedom = 162, the ratio of Chi-square to degrees of freedom = 1.70, RMSEA = 0.062, GFI = 0.87, NFI = 0.94, IFI = 0.98, TLI = 0.97 and CFI = 0.97, indicated a good model fit for this model and suggested that the structural relationships among the factors of two questionnaires in senior elementary school teachers was interpreted properly in this model. According to Figure 2, senior elementary school teachers’ GK was not only the significantly positive factor for explaining the variation in their GCK, GPK, and Attitude, but also the significantly negative factor for explaining the variation in their Actual teaching usage. Furthermore, senior elementary school teachers’ GPK was the significantly positive factor predicting both of their GPCK and Actual teaching usage.

Figure 1. The junior teachers’ structural model of their TPACK-G and acceptance of GBL.

Figure 2. The senior teachers’ structural model of their TPACK-G and acceptance of GBL.

4. Discussion

Findings of this study include, first of all, the insignificant difference between the junior’s and senior’s GPK might imply that the elementary schoolteachers hardly identify the difference between GPK and GPCK. Second, the junior teachers’ GK, GCK and GPCK outperformed the seniors, which is resonant to the prior research (Hsu et al., 2017). Third, GPK is the most predictive factor to their GPCK. Finally, the senior teachers’ GPCK is unable to predict their attitude and actual teaching usage, and their GK is negatively related to their Actual teaching usage. It was possible that the senior teachers’ GPCK is not developed; thus, they were less likely to know how games can be utilized to enhance the quality and effectiveness of their teaching practice.

References

Game-based Narrative System for Student English Learning

Zhi-Hong CHEN*, Huei-Jhen CIOU & Pei-Yun CHI

*Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan
bDepartment of Information Communication, Yuan-Ze University, Taiwan
*zhchen@ntnu.edu.tw

Abstract: In this paper we describe a game-enhanced language learning system that uses narrative as an interactive environment to anchor student language learning. The system consists of three layers, including game world, storylines, and dialogues. In addition to the development of the system, a preliminary system evaluation was also conducted. The results revealed that the system make students have a significant improvement in terms of their learning effectiveness.

Keywords: Game-based learning, English vocabulary, Narrative

1. Introduction

According to the principle of situated learning theory (Lave & Wenger, 1991), learning is advocated to be situated in a specific context, or embedded in a particular social and physical environment (Kindley, 2002). Based on the foundation, different approaches are investigated to explore how to apply this principle in practice: scenario-based learning (Clark, 2009), and game-based learning (Prensky, 2008). First, scenario-based learning places students in authentic works in order to integrate needed knowledge and skills situated in the context (Clarke, & Mayer, 2011). In other words, scenario-based learning advocates that students should learn in concrete situations and by examples, instead of abstract or decontextualized knowledge. In addition, game-based learning recently provides promising potentials to support situated learning through characteristics of digital games in student motivation and participation (Schultz & Fisher, 1988).

On the other hand, it has been indicated that vocabulary plays a significant role in mastering a foreign language (Nguyen & Khuat, 2005), and language learning and acquisition is better to occur from meaningful contexts, not separated from the learning situations (Krashen, 1981). Following this line of thought, different studies of game-enhanced language learning have been explored. This is due to the fact that digital games could stimulate student participatory motivation and improve learning effectiveness. For instance, Chen and Yang (2013) used digital games to help students acquire second language, and the results indicated that digital games could stimulate student motivation and further foster their reading, listening, and vocabulary skills. In addition, Chen and Tsai (2009) used mobile games to help students learn English, and the finding indicated that this game-based system could increase students’ interest and willingness to learn.

However, game-enhanced language learning often focuses on the game types and rule-based mechanisms, and ignores the potentials of narrative and dialogues. Previous study has indicated that character design and narrative environment are two features that could foster student learning (Dickey, 2007). Thus, there is a need to investigate how to integrate narrative with the development of game-enhanced language learning systems.

2. Interactive 3D game: PlanetAdventure

This study develops an interactive narrative game, named PlanetAdventure system, which extends the background setting of the novel “little prince” (de Saint-Exupéry, 2013). Students play the role of little
prince to find his lovely rose in the B612 planet, but the rose is murdered. Thus, the game goal is to expose the mystery: who murder the rose and the reason why. The PlanetAdventure system consists of three major components: game world, storylines and dialogues (see Figure 1).

2.1 Game World

The game world is established based on the background story of the novel “little prince”, which consists of several scenes in the planets. Each scene involves different non-player characters (NPCs) that can interact with students to trigger specific events, such as discovering cues, initiating dialogues, or starting specific events. The purpose of these different scenes in the planets is to offer students immersive contexts to learn, because they have more controls in the game world to decide where to explore, whom to talk, what object to pick up. These choices and controls in the game world allow them to create their own “story” and learning experiences.

2.2 Storylines

The storylines refer to the series of events that happen in the game world, and are also called “plots”. The main storylines of the PlanetAdventure system is to expose the murders who kill the rose, and contains three storylines: king “penguin”, conceited “parrot”, and lamplighter “fish”. Different from the original novel, these NPCs in this system are illustrated in the forms of cartoon animals to engage students in a fantasy world. Storylines are the main topics of the learning materials. When students advance in one of the storylines, they are situated in a specific scene. In this way, the entire game environment is learning materials, were students can interact with these NPCs and learn from the dialogues.

2.3 Dialogues

When students advance in one of storylines, they will encounter a number of NPCs, which will initiate pre-defined dialogues, and offer students information and cues about the murderer. Because the game goal is to figure out the mystery, they are required to interact with different NPCs for obtaining news and information, and then interpret and reason the cause-and-effect relationship. This purpose will drive students to understand comprehensively the information what the dialogues provide. In other words, such game environments could be regarded as “anchor” to combine learning materials, and students can learn from the sentences, vocabularies of the dialogues.

Figure 1. Screenshot of the PlanetAdventure system
3. Pilot Study

After the PlanetAdventure system was implemented, a pilot study was further conducted to preliminarily evaluate the influences of the system. The participants were 61 college students who used the system for 40 minutes to learn 30 specific English vocabularies. These English vocabularies were embedded into the dialogues of the system. Pre-test and post-test were employed before and after the system use, respectively. Both of the tests were the same, but they were formatted in different order to prevent the rote effect. The test contained was 30 blank-filling questions with the scores ranging from 0 to 100. The results of the pre-test and post-test were illustrated in Table 1. A paired-sample t-test was further conducted to examine its difference, and the finding indicated that the improved scores had a statistical significance, implying that the PlanetAdventure system make student improve their learning effectiveness. However, more future work is required to investigate the influences of the system, and obtain more feedback to revise the system development.

Table 1: Results of the pre-test and post-test

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th></th>
<th>Post-test</th>
<th></th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>10.73</td>
<td>SD</td>
<td>5.12</td>
<td>25.14</td>
<td>4.74</td>
</tr>
</tbody>
</table>

Acknowledgements

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for financial support (MOST-105-2511-S-003-051).

References


Exploring Teachers’ Pedagogical Design Thinking in Game-based Learning

Mingfong JAN*, Wan-Lin YANG & Michael THOMAS
Graduate Institute of Learning and Instruction, National Central University, Taiwan
Center for Teacher Education & Institute of Education National Cheng Kung University, Taiwan
Educational Psychology University of Illinois at Chicago, USA
mingfongjan@gmail.com

Abstract: In this paper, we investigate teachers’ pedagogical design thinking while enacting game-based learning in a primary school. We employ a qualitative case study approach to study three cases that reveal teachers’ pedagogical design thinking. Three emerging themes regarding teachers’ game-based learning fallacies are identified through the above three cases: (1) misinterpreting game affordances, (2) viewing a role-playing simulation game as a resource for content knowledge, (3) employing direct instruction as a dominant GBL pedagogy. This study unpacks how teachers think while they enact game-based learning, which is currently under-researched in the game-based learning literature.

Keywords: game-based learning, GBL, GBL enactment, game-based pedagogy

1. Introduction

Since its emergence more than 15 years ago (Gee, 2003; Prensky, 2003), game-based learning (will be referred to as GBL hereafter) has been framed as more of a technological innovation than a pedagogical innovation (Jan & Gaydos, 2016). Research on GBL predominantly foregrounds how games, digital and non-digital, can be designed to foster learning (e.g., Kiili, 2005; Papastergiou, 2009). It is not until recently that researchers pay attention to how teachers interpret and enact games with learning activities (Jan & Gaydos, 2016).

This qualitative case study aims at developing an in-depth understanding of how teachers enact GBL in a primary school that specializes in GBL for seven years. Specially, we identify the issues that fifth grader science teachers face in bringing GBL to their classrooms. Three separate studies are conducted: (1) a Professional Development (PD) workshop designed to fostering GBL enactment, (2) two full cycles of GBL enactment using the Prey and Predator role-playing game with two science teachers, and (3) a focus group discussion with six Primary Five science teachers who use the Prey and Predator game and have used other games in their classrooms. Research questions are:

1. How are teachers trained to enact game-based learning?
2. How do teachers enact game-based learning?
3. What are the emerging pedagogical issues associated with game-based learning enactment?

2. Methodology

We employ a qualitative case study approach to understand the above research questions. Unlike quantitative approaches which are often employed to produce frequency counts or verify existed models (Maxwell, 2012), case study enables us to investigate the evolving processes through which teachers conceptualize and enact GBL by drawing on multiple sources of data.

Participants Selection: Participants are ten teachers Primary 5 (fifth grade) teachers from a Singapore primary school. Most of them are science teachers while some teach language arts, math, or mother tongue. All of them have used games in their classrooms for at least one year. Two of them are subject heads who have more than five years teaching experience.
School Selection: The selected school is noted for its involvement, long-term development, and dedication to GBL. It is one of the first primary schools to send teachers to participate overseas GBL workshops. At the time of the study, the school strived to become a signature school for GBL.

Prey and Predator Game: The Prey and Predator role-playing game is co-developed by the school and a GBL vendor. Students role-play as either Preys or Predators in teams. The goal is to survive either by escaping predators or catching preys.

Cases and data collection: We collect data from three separate cases: (1) a 2-hour professional development workshop designed to help teachers understand the Prey and Predator role-playing game; (2) two full cycle GBL program using the Prey and Predator game; (3) a 2-hour focus group discussion with all participant teachers. Data are collected from (1) observation and field notes of the above cases, (2) transcripts of interviews, (3) teachers’ lesson plans.

Data analysis: All collected data are open-coded, categorized, and thematized with qualitative coding techniques (Glesne & Peshkin, 1999; Maxwell, 2012). We triangulate (Glesne & Peshkin, 1999) multiple sources of data while identifying emerging and common themes. This improves the trustworthiness of our interpretation.

3. Findings

3.1 Case One: Professional Development (PD) workshop designed to foster GBL enactment

Conceptualized by the future school, the workshop is to inform teachers how to use the Prey and Predator game and how to win the Prey and Predator game. A technician conducted the Professional Development workshop in the computer lab. The instruction focused on teaching technical know-hows as well as strategies for winning. The strategies for winning became crucial learning activities when P5 teachers enacted Prey and Predator later on. In the professional development workshop, the followings are raised, but never discussed—issues regarding the design of learning activities, how students learn, and what students learn with the Prey and Predator game. In a nutshell, the PD session engaged teachers in thinking about game-related technical issues. Though there were pedagogical issues being raised, none were discussed among teachers.

3.2 Case Two: Two full cycle GBL enactment using Prey and Predator, a role-playing game that the school develops with a vendor. 7 interviews with 2 teachers who enacted the GBL learning program

The Prey and Predator learning program is co-designed by all Primary 5 science teachers as a part of an integrated program. The learning program is composed of (1) the Prey and Predator role-playing game and (2) learning activities that connect the game to textbooks and curricular standards. It is designed for learning the concept of adaptation and for developing collaboration skills. As all P5 science teachers design the learning program together, the enactments demonstrate their collective design thinking.

In the enactment, teachers used the Prey and Predator game as a simulation of the African Safari (and it is far from a real world simulation), asking students to “apply” biological and structural adaptation concepts when they played the Prey and Predator game. They ask students to play as a team all the time as forming a team is viewed as a sign of collaborative learning. As the game was far from a real world simulation, their pedagogical approach did not match the designed affordances of the game. It was also impossible for students to collaborate as teams as the teachers prohibited dialogues. In conclusion, the two teachers, both having taught the course for a few years, demonstrated questionable design thinking. First, they were not able to understand and leverage the affordances of the Prey and Predator game. Second, they employed a direct instruction approach to teach collaboration.
3.3 Case Three: Focus Group Discussion (FGD): The history of GBL in the school and the issues they have experienced in this process

The FGD is designed to understand teachers’ experience with GBL and their challenges in general. The preliminary analysis suggests that

1. Teachers spontaneously used two kinds of games in their classroom—games that motivate learning and games that provide drill and practice.
2. No teachers have ever used games that foster 21st century competencies because they were not confident in using/designing them at all.
3. There were several workshops designed to foster technical know-how, and they were the only professional development sessions they had about GBL.

4. Discussion – Teachers’ Fallacies regarding GBL

Three emerging themes regarding teachers’ GBL fallacies are identified through the above three cases: (1) misinterpreting game affordances, (2) viewing a role-playing simulation game as a resource for content knowledge, (3) employing direct instruction as a dominant GBL pedagogy.

The Prey and Predator game is a role-playing game that affords players to play either as preys or predators. They survival of the preys or predators relies on gaming strategies, which have little or nothing to do with natural laws. Therefore, it should not be used as a “simulation” of African Safari. Asking students to “apply” biological and structural adaptation in playing the Prey and Predator game results in a mismatch between the designed affordances and perceived affordances of the game.

The above issue is actually embedded in a larger context of pedagogy. The two teachers enacting the Prey and Predator game mostly employ direct instruction in their everyday teaching. In direct instruction, the teachers give and explain answers to their students. Their pedagogical approach is a noticeable mismatch with the Prey and Predator game because there are no teachable contents in the game. As a result, teachers add contents into the game from textbooks. In fact, the teachers not only teach biological and structural adaptation, ask students to apply the concepts, but also “telling” students how to win the game when students failed.

Our studies conclude with an alarming sign for teaching with innovative technology – the development of teachers’ expertise. Although games and other learning technologies have been claimed to be innovative for learning, we have yet innovated teaching. Our teachers have done their best; our teacher education systems have just begun to explore such issues.

Acknowledgements

We would like to thank the Ministry of Science and Technology, Taiwan, and National Institute of Education, Singapore, for funding the research and conference presentation under project number 105-2511-S-008-011-MY2.

References

Can Conversational Agents Foster Learners’ Willingness To Communicate in a Second Language? : Effects of Communication Strategies and Affective Backchannels

Emmanuel AYEDOUN*, Yuki HAYASHI & Kazuhisa SETA
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan
*eayedoun@ksm.kis.osakafu-u.ac.jp

Abstract: Willingness to communicate (WTC) in a second language (L2) is believed to have a direct and sustained influence on learners’ actual usage frequency of the targeted language. To help overcome the lack of suitable environments to increase L2 learners’ WTC, our approach is to build conversational agents that can help L2 learners overcome their apprehension towards communication in L2. In this paper, we focus on the dialogue management aspects of our approach and propose a model based on a set of communication strategies (CS) and affective backchannels (AB) in order to foster agents’ ability to carry on natural and WTC friendly conversations with L2 learners. An evaluation of the proposed method led to two main findings. First, combining CS and AB empowers the conversational agent, making possible highly significant WTC gains among L2 learners in English as a foreign language context. Secondly, even a single implementation of AB proved to have the potential to enhance L2 learners’ WTC to some extent.

Keywords: Willingness to communicate in L2, conversational agents, communication strategies, affective backchannels, intelligent tutoring

1. Introduction

One of the fundamental goals of second language (L2) learning is to provide learners with the ability to communicate effectively using their L2 when given the opportunity to do so. The key factor to ensure such communicative readiness is the willingness to communicate (WTC) defined as the “readiness to enter into discourse at a particular time with a specific person or persons, using an L2” (MacIntyre, et al., 1998). Following the finding that learners with a higher WTC tend to perform better than others in producing the target language, MacIntyre and his colleagues suggested that increasing learners’ WTC should be the ultimate goal of L2 learning. Moreover, they proposed a pyramidal heuristic model of variables affecting WTC in which it appears that the environment where learners experience or practice the L2 plays an important role in motivating them to actively take part or not in L2 conversation. However, as pointed by Reinders and Wattana (2014), many learners feel a genuine anxiety about performing in front of others, and many classrooms do not, as a result, offer learners much in the way of communicative practice as would be desirable.

The ultimate goal of this study is to contribute in enhancing L2 learners’ WTC by providing them opportunities to freely simulate and enjoy immersive daily conversations using a computer-based conversational environment. However, L2 communication is also problematic because it involves learners’ ability to communicate within restrictions on their own vocabulary, grammar, etc. Thus, unlike communication between L1 learners, breakdowns or pitfalls in communication occur more often here. Therefore, any conversational agent intended to support communication in L2 should adopt some strategies adapted to such interactions.

In this paper, we propose and evaluate a dialogue management model, based on a set of specific conversational strategies, namely Communication Strategies (CS) and Affective Backchannels (AB), in order to foster dialogue agents’ ability to carry on WTC effective conversations with learners in English as a Foreign Language (EFL) context.
2. Related Works

Some L2 learners despite excellent linguistic competence tend to avoid using L2 for communication where as some others with only minimal linguistic competence seem to communicate in the L2 whenever possible. MacIntyre et al., explained such differences arguing that the intention or willingness to engage in L2 communication, rather than linguistic competence, is determined by a combination of immediate precursors such as learners’ perception of their own second language proficiency (perceived competence), the opportunity to use the language (context), and a lack of apprehension about speaking (L2 anxiety). Following these findings, many researchers from different countries, such as Yashima in Japan (Yashima, 2002), Peng in China (Peng, 2007), or Oz in Turkey (Oz et al., 2015) have intensively investigated the validity of the WTC model in their own respective contexts. Although some differences, mainly due to each country’s cultural and social specificities, may exist, it’s generally admitted that the variables identified by MacIntyre and his colleagues can be seen as a basic and universal reference model of key factors influencing WTC in L2. Furthermore, WTC studies have shown that learners displaying high WTC are more likely to show more improvement in their communication skills (Yashima et al., 2004) and to acquire higher levels of language fluency (Derwing et al., 2008).

However, it is interesting that even though research investigating computer-mediated communication in the context of second language acquisition (SLA) has proliferated since more than two decades now, only few of them have actually investigated or proposed practical ways to enhance L2 learners’ WTC. Compton (2004), for example, revealed that chatting helped students to feel confident and consequently, willing to participate orally in class discussions. However, its impact on WTC varied from learner to learner and was dependent on a number of factors, particularly the topics of discussion and the attitudes of their partners. Nakaya and Murota (2013) developed a mobile conversation learning system, which aimed to motivate Japanese EFL learners. In their system, conversation topics were based on learners’ life logs or related to situations that learners often experience in their daily life. Still, the “conversations” were mainly system-driven so that learners were just limited to answer questions generated without any possibility for them to get help from the system when they face difficulty in answering some questions. But difficulties and breakdowns do occur in L2 conversation. In our previous work, we proposed an embodied conversational agent based on MacIntyre’s WTC model to help increase L2 learners’ WTC by providing learners with opportunities to naturally simulate daily conversations in various social contexts (Ayedoun et al., 2016). Our evaluation of the system demonstrated its potential to simulate efficiently natural conversations in a specific context as well as the feasibility of improving learners’ WTC using such a computer-based environment. Nevertheless, conversation opportunities alone are not enough to effectively motivate learners towards communication in L2. That is why we mentioned that a good level of conversation smoothness, to be achieved by implementing strategies to keep the conversation going on especially when learners face some difficulties, are desirable since they contribute to creating a friendly conversational environment and reduce learner’s anxiety.

As pointed out by Mesgarshahr and Abdollahzadeh (2014), language learners, especially at lower levels, are highly likely to experience some difficulty when communicating in the target language. They added that too much difficulty during communication might make them abort their attempt to carry on and, consequently result in dissipating the desire to communicate. Thus, being able to help learners overcome difficulties when communicating in L2 can be considered as an essential ability for any conversational agent intending to increase WTC among L2 learners. Communicative strategies known as CS are defined as “a systematic technique employed by a speaker to express his [or her] meaning when faced with some difficulty”. For example, approximation is a CS which consists in using a term that expresses the meaning of the target lexical item as closely as possible (e.g. “the thing you open bottles with” for “corkscrew”). CS have been an object of intense research in the area of SLA, Dörnyei and Thurrell (1991) referring to this ability as strategic competence. Strategic competence is considered by Canale and Swain (1980) as one component of communicative competence. It is conceivable that an underdevelopment of this competence may account for some L2 learners’ lack of the ability to overcome interactional pitfalls, which may adversely affect their WTC.

As indicated by Mesgarshahr and Abdollahzadeh, in the case of learners with a low WTC, achievement of communicative competence does not automatically guarantee L2 usage. This is because L2 WTC is also directly affected by affective variables such as anxiety or self-confidence, which
contribute to determine the psychological preparedness to communicate at a particular moment. It might therefore also be necessary for the dialogue partner (for instance the conversational agent) to provide frequent dialogue feedbacks, also known as backchannels, which reassure the learner and makes him willing to pursue the interaction. Previous works by Kopp et al., (2007) as well as Morency et al., (2010) have amply demonstrated the importance of such backchannels in human-agent conversation but here, we are interested in a specific category of backchannels that we call “affective backchannels” (AB). We believe that they might provide an effective support to WTC affective variables in a computer-mediated interaction with L2 learners.

3. Conversational Strategies to Increase WTC

3.1 Contribution and Novelty

When it comes to propose practical ways to make learners more willing to communicate in L2, most of the significant contributions have been conducted in the fields of communication studies or language learning studies. In the fields of computer assisted language learning (CALL) and artificial intelligence in Education (AIED), the topic seems to be a conspicuous rarity in the literature since traditional spoken dialogue frameworks seem to not particularly take into consideration aspects related to L2 learners’ WTC. Besides, among the few research studies dedicated to propose a computer-based approach to increase levels of L2 WTC, less effort has been expended on investigating usage of virtual realistic interfaces such as embodied conversational agents, which yet seem to have the potential to be an efficient alternative to real interactions. Following our previous work (Ayedoun et al., 2016) in which we showed that a dialogue agent based conversational environment might be useful to increase L2 learners’ WTC, we propose a dialogue management model dedicated to facilitate the implementation of intelligent conversational agents that are effective in increasing L2 learners’ WTC. The originality of our approach lies in the fact that the proposed model takes into consideration both aspects related to communicative breakdowns that occurs very often in L2 learners-agent interactions and those related to affective variables influencing L2 WTC according to MacIntyre’s WTC model.

Many studies in various fields such as education and psychology have contributed to proposing several theories regarding teaching and learning processes between human teachers and learners. Although researchers such as Nass, Steuer and Tauber (1994) have amply demonstrated that the human-computer relationship is fundamentally social, the above mentioned teaching and learning theories may not be similarly applicable in agent-human learning situations, for the simple reason that it is still not clear whether all learners react similarly to computer agents, as they would do with human partners. Thus, through this research, we aim not only to contribute to enhancing L2 learners with a computer agent based system but also whereby to collect quantitative and qualitative data about the relationship between conversational agents and L2 learners, which might be useful to propose a generic model of the characteristics of such interactions (long term goal).

3.2 Proposed Dialogue Management Model

The model aims first, by the way of Communication Strategies (CS) to foster the dialogue agent’s ability to autonomously detect and robustly handle recognition errors as well as learners’ pitfalls in L2 communication, making possible achievement of more or less smooth interaction between L2 learners and dialogue agent. Secondly, by the way of Affective Backchannels (AB), this model aims to make possible achievement of a warm interaction where learners feel less anxious about L2 communication and progressively get confidence about their own linguistic proficiency.

Communication Strategies (CS): CS are “a systematic technique employed by a speaker to express his or her meaning when faced with some difficulty” (Dörnyei and Scott, 1997). These difficulties might arise either from the speaker (lack of linguistic resources) or from the interlocutor (impossibility to understand the speaker). It is worthwhile for learners to have a repertoire of such strategies at their disposal, whereby they achieve a degree of communicative effectiveness beyond their immediate linguistic means (Thornbury, 2005). Nevertheless, in the case of learners with a low WTC, mastering such strategies does not necessarily guarantee that learners will be able to use them when they face
some trouble during conversation. On contrary, the use of CS might help dialogue agents not only to overcome their own difficulties (impossibility to understand the learner…) but also and more importantly to anticipate or handle more effectively communication pitfalls (difficulty in understanding or answering) that learners may encounter during conversations. When learners know that they can rely on a supportive dialogue agent to help them recover from difficulties, they may feel a “sense of security” that can reduce their communication apprehension, leading to a higher level of WTC. In the present study, we targeted about 9 strategies among those defined in the comprehensive review of definitions and taxonomies of CS (Dörnyei and Scott, 1997). The selected strategies were chosen according to two criteria: (i) their effectiveness towards encouraging WTC and (ii) the feasibility of their implementation from the technical standpoint. Table 1 shows a non-exhaustive list of the selected strategies as well as examples of their usage in this study.

Table 1: Example of CS implemented in this study.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplification or Approximation</td>
<td>Use an alternative or a shorter term, which expresses the meaning of the target lexical item.</td>
<td>Agent: May I have your order please? Learner: … (silent) Agent: Order please</td>
</tr>
<tr>
<td>Code switching</td>
<td>Use a L1 word with L1 pronunciation or a L3 word with L3 pronunciation in L2.</td>
<td>Agent: May I have your order please? Learner: … (silent) Agent: Go chūmon wa ikagadesu ka? (Code switching from English to Japanese)</td>
</tr>
<tr>
<td>Ask clarification</td>
<td>Request explanation of an unfamiliar meaning structure.</td>
<td>Learner: One xxx please. Agent: What do you mean?</td>
</tr>
<tr>
<td>Suggest AP (Answer Pattern)</td>
<td>Provide an example of answer that could fit the current discourse context</td>
<td>Agent: What would you like to drink? Learner: … Agent: For example, you may say one beer please to order a beer.</td>
</tr>
</tbody>
</table>

Affective Backchannels (AB): Backchannels are generally defined as a type of short utterances or feedbacks such as uh-huh, yeah… given by the listener to show interest, attention or a willingness to keep the communication channel open. They play an important role in human agent conversation [Smith 11]. Although actual competence might encourage communication, it is the perception of that competence that will ultimately determine the choice of whether to communicate or not (Clément, Baker and MacIntyre, 2003). Thus, L2 learners who don’t get enough supportive feedbacks from their interlocutors may perceived themselves as being incompetent communicators and therefore tend to be reticent to communication. All this gives much evidence that it might be effective for a conversational agent intending to enhance learners’ WTC, to be able to convey a sufficient amount of interest or sympathy to learners during the interaction since doing so might contribute to creating a WTC friendly atmosphere. In order to achieve such empathetic support, we identified and defined a set of backchannels that we call Affective Backchannels (AB). Table 2 shows the different categories of AB that we defined in order to cover a wide range of situations the learner may experience during the interaction.

Table 2: Implemented AB in this study.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congratulatory AB</td>
<td>Employed when the conversation with the learner is going well as expected</td>
<td>Okay, that’s nice!</td>
</tr>
<tr>
<td>Encouraging AB</td>
<td>Employed when the learner seems to hesitate to the extent that he/she remains silent.</td>
<td>Come on, you can do it!</td>
</tr>
<tr>
<td>Sympathetic AB</td>
<td>Employed when the learner’s utterance does not match the agent expectations.</td>
<td>Sorry I couldn’t get you…</td>
</tr>
<tr>
<td>Reassuring AB</td>
<td>Employed when the learner seems to face much difficulties in the conversation.</td>
<td>Don’t worry…</td>
</tr>
</tbody>
</table>

The core architecture of our conversational agent was developed in our previous work (Ayedoun et al., 2016) and is composed of two main components (the dialogue manager and the multimodal response generator) connected to several external web-services and resources as shown in figure 1(top). The overall conversational flow is under the supervision of the dialogue manager, which controls the various phases of dialogue and their timing, as well as the level of system initiative, in an integrated fashion. As described in figure 1 (bottom), the dialogue strategies management routine goes from Start to End (top to bottom of the figure) passing through checking of the different possible dialogues states represented in the diamond symbols. The occurrence of each of such dialogue states automatically leads to triggering of adapted conversational strategies (as indicated in square symbols) that are pull out from their respective databases (as indicated by dotted lines) in order to keep the learner motivated using AB (represented in pink color database symbols), and try to move the dialogue forward using CS (represented in blue color database symbols). The decision to engage a specific conversational strategy is mainly based on the following triggering events or dialogue states:

Figure 1. System architecture showing interface (top), Dialogue management model based on AB and CS (bottom)
The learner is silent: when the system is expecting some input from the learner but cannot get any after a certain amount of time is elapsed. In such case, the system will first apply a Reassuring or Encouraging AB and then investigates the reason why the learner remains silent by asking appropriate questions.

The learner is NUNA (Not able to Understand, Nor to Answer): when the learner is not able to understand what the agent is expecting from him. In such case, the system will fire up specific CS such as Simplification in order to let the learner understand and hopefully utter the expected information.

The learner is UNA (able to Understand but Not able to Answer): when the learner understands what is being requested from him but can’t or don’t know how to answer. In this case, CS such as Suggest an Answer Pattern will be applied in order to give the learner a hint about how to overcome his current difficulty.

The learner is asking for help: when the learner expresses that he is NUNA, UNA or specifically requests a CS such as repetition or simplification. In this case, the system will fire up a Reassuring AB and then apply appropriate CS according to the nature of the help requested by the learner.

The agent is NUNA: when the system is unable to detect the learner’s intention due to a very low confidence score or the occurrence of a recognition error in the processing of the learner’s utterance. In this case, the system will first output a Sympathetic AB and then try to recover by applying CS such as Ask repetition in order to give the learner another chance to express his intention.

The agent is UNA: when the system is able to detect the learner’s intention with an acceptable confidence rate but is not expecting such intention in the current dialogue context (for example, the learner asking for the nearest supermarket while the agent is expecting him to make an order in a restaurant context). In this case, the agent will first apply a Sympathetic AB and then try to get the learner reformulate his intention by using CS such as Ask confirmation in order to make sure that what the system understood from the learner’s utterance is actually what the latter actually meant.

We expect that the modular and domain independent nature of the proposed dialogue management model will not only facilitate its reusability across different dialogues domains, but will also make easier the development of conversational spoken languages interfaces that are more adapted to L2 learners from the WTC standpoint.

Figure 2 shows an excerpt of conversation between a learner and the agent depicting how AB (in pink) and CS (in blue) are called into action according to the different dialogue states (in gray). As shown, the successive interventions of the system are successful in gradually help the learner overcome his initial breakdown following the question “Would you prefer smoking or non smoking table?” As illustrated, it is the successful detection of the occurring pitfalls combined with the help provided by the agent through usage of appropriate AB and CS that ultimately led to the conceivable answer “Non-smoking please” from the learner. Without the use of such strategies, the conversation would have probably ended up just after the agent’s first question since the learner seemed to be unable to go further into the interaction. It is this kind of support that our dialogue agent aims to provide to L2 learners through usage of AB and CS.

Agent: Good evening. Do you have a reservation?
Learner: No.
Agent: Ah, I see! [AB: Congratulatory]
Would you prefer smoking or non smoking table?
Learner: ... (“Learner is silent”)
Agent: Come on, don’t give up! [AB: Encouraging]
Do you understand?? [Comprehension check]
Learner: No . (“Learner is NUNA”)
Agent: Which area do you prefer.
[CS_L_NUNA: Simplification]
Learner: I don’t know area. (“Learner is UNA”)
Agent: Don’t worry, I’m here to help you.
[AB: Reassuring]
For example, you could say: “Non-smoking area please!” [CS_L_UNA: Suggest AP]
Learner: xxx xxx please! (“Agent is NUNA”)
Agent: Sorry, I’m afraid I missed something.
[AB: Sympathetic]
Could you please repeat that?
[CS_L_NUNA: Ask repetition]
Learner: Non-smoking please.
Agent: Great! [AB: Congratulatory]
Please come this way…

Figure 2. Excerpt of actual dialogue between the agent and a learner illustrating use of AB and CS.
5. Experimental Study

We conducted an evaluation of the proposed dialogue management model in this paper to clarify the following preoccupation: Does the usage of CS and AB really have the potential to empower the conversational agent to the extent to foster L2 learners’ WTC?

5.1 Procedures and Materials

**Conversational agent:** We employed a conversational agent based on the system architecture proposed in our previous work (Ayedoun et al., 2016), and enhanced it with the management model described above. The system makes possible spoken dialogues between the conversational agent personified as Jack, on one hand and learners on the other.

**Participants:** The study was conducted with 40 Japanese undergraduate and graduate students currently attending a Japanese university. In terms of language background, participants were fairly homogenous; all of them were native Japanese speakers and none had experience of living in an English-speaking country. They were informed that their participation to the study was voluntary and the results would be anonymised.

The evaluation was conducted following 5 procedures as described in Table 3.

**Procedure 1 and Procedure 3 (Measures of WTC):** We employed a widely used survey developed by Matsuoka (2006) and inspired from Sick and Nagasaka’s WTC test (2000) to evaluate learners’ WTC before (procedure 1) and after (procedure 3) they interacted with the system in procedure 2. It is worth to mention here that we privileged such a self-report estimation of learners’ WTC before and after the interactions because it is the common approach used in related works to measure L2 learners’ WTC. Moreover, the literature teaches us that it is learners’ perception of their own competence that will ultimately determine their choice of whether to communicate or not, so that inquiring directly learners about their WTC might actually give a good indication about their future behavior in actual communication situations. The WTC questionnaires targeted three variables: confidence, anxiety and desire to communicate considered as the immediate precursors of WTC (MacIntyre and Charos, 1996). In the WTC questionnaires, participants were asked to rate each of these variables in 30 scenarios (e.g., Making a telephone call in order to make a reservation at a hotel in an English speaking country) related to using English in various circumstances with a four-point Likert scale. All participants were given as much time as required to complete the questionnaires. Data were collected anonymously via an online survey service and participants were told that their answers would be kept confidential. In order to minimize order effects, we carefully designed learners’ interactions with the system in each group applying the counterbalancing method proposed by Howitt and Cramer (2011).

**Procedure 2 (First Interaction with Jack):** All the participants were initially asked to interact with Jack, who would teach them how to pronounce some words in English. They were requested to listen and repeat the words according to Jack’s instructions. In reality, our intention here was to let all the learners sympathize with Jack and understand how the system works. Then, participants were split into four groups (Group 1 to Group 4) of 10 participants each, making sure that there were not statistically significant WTC differences in the pretest results among all the four groups. Doing so ensure an homogeneity in the starting conditions, which suggests the validity of comparing WTC gains among the different groups after the interactions with the agent. Participants in each group were then asked to interact with the system, the conversation being held this time in a restaurant context with Jack interacting with them as a waiter. We prepared 3 different versions of the system: the CS+AB version (with both CS and AB implemented), the CS version (with only CS implemented) and finally the AB version (with only AB implemented). Participants interacted with a version of the system according to their group. For example, participants in Group1 interacted with the CS+AB version, those of Group 3...
with the CS version and so on, as indicated in Table 3. It is also important to mention that participants interacted individually with the system in a room specially prepared for the evaluation and were given as much time as they wish to enjoy the conversation with Jack until they were requested to pay the bill. They were also informed that they were free to interrupt the interaction at anytime in case they feel to do so.

**Procedure 4 (Second Interaction with Jack):** After taking the second WTC questionnaire (posttest) in procedure 3, participants were asked again to interact with the system in a restaurant context. As in Procedure 2, participants interacted with different versions of the system according to their groups but were not informed that the system is different from the one they used in their first interaction.

**Procedure 5 (System preference survey):** After procedure 4 described above, all participants were asked to choose which one of the two interactions (i.e.: which version of the system) they preferred the most as well as the reason supporting their choice. For example, participants in Group 1 had actually to choose between the CS+AB and the CS version, those of Group 2 between the CS+AB and the AB version, and so on for participants in Group 3 and Group 4.

From the WTC standpoint, we assume that the results would be viewed as positive if the interactions with the conversational agent led to improving participants’ confidence and desire to communicate while reducing their anxiety.

### 5.2 Results

Figure 2 shows variations observed among participants of each group before and after interacting with the system in terms of WTC variables (confidence, anxiety, desire to communicate). These results indicate that WTC gains (increase in confidence and desire to communicate with decrease in anxiety) were observed across all groups, irrespectively of the system version (CS+AB, CS or AB). Then, we performed paired samples t-tests and measures of effect size (using Cohen’s $d$) to evaluate how significant were the WTC gains in each group.

There were statistically highly significant differences between the first and the second WTC questionnaires administrated to participants in Group 1 and Group 2 after they interacted with the CS+AB version. Actually, their confidence and desire to communicate increased respectively by +0.60 ($t(9) = -8.91, p<.001, d=2.3$) and +0.63 ($t(9) = -5.80, p<.001, d=1.7$) in group 1, by +0.70 ($t(9) = -4.53, p<.01, d=1.6$) and +0.64 ($t(9) = -8.05, p<.001, d=1.3$) in group 2, while their anxiety decreased by -0.64 ($t(9) = -4.42, p<.01, d=1.6$) in group 1 and by -0.82 ($t(9) = -6.63, p<.001, d=1.8$) in group 2.

There were not statistically significant WTC gains among participants in Group 3.
There were statistically significant differences between the two WTC questionnaires among learners in Group 4 after they interacted with AB version. Actually, their confidence and desire to communicate increased respectively by +0.50 \( [t(9)=-2.82, p<.1, d=1.6] \) and +0.33 \( [t(9)=-2.35, p<.1, d=0.6] \) while their anxiety decreased by -0.58 \( [t(9)=3.48, p<.01, d=1.2] \).

Besides, the preference rate of the CS+AB version was constantly high across all the 4 groups being preferred by 32 participants out of 40 (80%) in total, while the CS and AB version have been preferred respectively by 5 participants out of 20 (25%) and 3 participants out of 20 (15%) as shown in figure 3. It is important to mention that this tendency has been observed across all the 4 groups no matter the order in which learners interacted with the CS+AB version. Actually, participants who preferred the CS+AB version, for example, frequently mentioned that they found natural and warm the way Jack showed some empathy throughout the interactions and appreciated the help they got from him when facing difficulties in understanding or expressing what they have got to say.

5.3 Discussion and Limitations

The results above allow us to draw a number of preliminary conclusions. Firstly, the combination of CS and AB proved to be really effective in motivating L2 learners, much more than just implementing CS or AB alone, as described above. This confirms our initial beliefs that making possible smooth and interactive conversations by using CS is not, by itself sufficient to increase effectively L2 learners’ WTC, which also requires the ability to convey a sufficient amount of warmness or sympathy to learners during the interaction via AB. The proposed dialogue management model in this paper covered both of these requirements and the results obtained are meaningful in terms of validating our approach. More interestingly, the results suggest that even a single implementation of AB could be quite effective in significantly reducing learners’ anxiety and contribute to increasing their WTC. This is a really interesting finding since it seems to reveal that the affective support that learners get during interactions might be the most important key factor influencing their WTC, so that a conversational agent able to provide a careful empathic support to learners might be quite beneficial for them.

Nevertheless, a possible limitation of this study is that the current version of our system is limited to conversations in only one context (restaurant context). We understand that learners’ WTC, of course, do not increase overnight and a certain amount of continuous usage of the current system with the possibility for learners to converse in various contexts is certainly necessary before we can collect more reliable data to support our findings, not to mention that we will certainly have to increase the sample size.

6. Conclusion and Future Research Directions

This paper has described a dialogue management model based on a set of two conversational strategies (CS and AB) aiming to empower conversational agents in order to foster L2 learners’ WTC in EFL context.

The evaluation results showed that the combination of CS and AB as proposed here is particularly effective considering the high WTC gains observed among participants who interacted with the CS+AB version of the system. We also found that even a single implementation of AB has the potential to enhance L2 learners’ WTC to a certain extent. Future research should be directed to confirm the tendencies evoked above by evaluating in more details effects associated with each strategy (CS or AB), determining approaches for strengthening their impact in enhancing L2 learners’ WTC, and carry out mid-long term evaluations about the outcomes of our approach on learners’ actual involvement in communication. We hope that this work will have genuine value and make a small contribution for proposing more effective computer based intelligent approaches to enhance WTC among L2 learners.

References


Multimodality in Language Education – Exploring the Boundaries of Digital Texts

Anna-Lena GODHEa* & Petra MAGNUSSONb
aUniversity of Gothenburg, Sweden
bUniversity of Kristianstad, Sweden
*anna-lena.godhe@gu.se

Abstract: Considering the challenges of contemporary communication, largely due to the rapid development of media technology, the purpose of this article is to discuss the implications that these challenges may have for language education. In a digitalized society, texts more often than not include several modes, such as images, symbols, film clips and sound. What does this widened notion of texts mean for language subjects and teachers and how are these changes in texts reflected in curricula? In the article we use a multimodal framework to understand and discuss meaning-making when a widened notion of texts is incorporated in language education and the conflicts that may occur are discussed from an activity theoretical point of view. We conclude by arguing for native language education to fully embrace and recognize multimodal meaning-making.

Keywords: Multimodal texts, language education, literacy

1. Introduction

Education today faces many challenges. In Sweden, the results of the PISA and PIRLS tests have revealed a declining trend which also includes reading comprehension. These tests have had a great impact on the political debate and have led to changes in curricula (Skolverket, 2011) with the aim to focus on basic skills and be more precise about standards concerning both teaching and learning. At the same time, technological advances means that meaning-making today involves texts which are increasingly multimodal, so that young people are socialized into a world where communication is accomplished by using a number of meaning-making modes (Statens Medieråd, 2015). In other words, young people read and create texts which are often a combination of written texts, pictures, moving images and sound mediated through digital devices such as computers and mobile phones. New methods of communication thrive in informal settings, whereas formal education often clings to traditional ways of making meaning, for example, when requiring written, paper-based texts from the students (e.g. Kress, 2010). At this moment in time, Swedish curricula are about to be changed again, this time the changes are made in order to strengthen students’ digital competence (SKOLFS, 2017:16).

Many young people today have constant access to texts in different modes and through different media. Widening the concept of text, to modes other than the verbal opens up other ways of meaning-making, and calls for new apt concepts to fit the communicative possibilities. Björkvall (2009) explains that ‘text’ in a multimodal perspective can be seen as different modes that may appear in various materialities. A verbal written text, for example, can be materialized as print on paper but also on a computer screen. But as texts, according to Björkvall (2009), are meaningful and coherent with a beginning and an end, it is difficult to talk about texts, based on this definition, when it comes to, for example blogs, websites and fan fiction. As a more apt overall concept, we use meaning offering, and consider it as being equivalent to available design (further explained and developed below). A meaning offering is connected both to the intended and the perceived meaning (Selander & Kress, 2010). As pointed out by Kress (2003), the development of technology, puts meaning-making in a new light. Other changes in society that needs to be taken into account include an increasing flexibility in the labor market, globalization, individualization, and increasingly multicultural societies (Cope & Kalantzis, 2010; Kalantzis & Cope, 2012). All of these changes affect teaching and learning in educational settings.
Despite the changing conditions for meaning-making, verbal communication in the shape of traditional reading and writing, continue to be necessary skills. They are crucial in society in order for citizens to take active part in democratic processes and citizenship. To acknowledge the way young people experience texts and, at the same time, support their reading and writing development, is a challenge for all teachers, but maybe in particular for native language teachers, i.e. teachers in the subject of Swedish in Sweden. In the syllabus and other documents that regulates the subject of Swedish, the development of students’ media and communication skills are recognized to some extent. The content of the subject has been expanded when incorporating movies, blogs and digital techniques for producing communication in the curricula. However, the revisions that are currently being made in the national Swedish curricula in order to strengthen digital competences in different subjects largely denote digital competence as concerning the use of digital tools and do not address changes in content to any great extent. This way of relating to changes arisen due to alterations in communication, is by Lankshear and Knobel (2008) considered to belong to an ‘old’ mindset. The use of digital devices for writing typographical texts can be taken as an example of the old mindset, where an established practice has been altered by the use of technology so that the writing of texts today is often done by using keyboard and screens, rather than pen and paper. Lankshear and Knobel (ibid.) consider an expanded concept of literacy to relate to a ‘new’ mindset, not to new as a measurement of time. This new mindset is understood as more collaborative and participatory. Knowledge is considered to be collective and distributed rather than centered on individual expertise. Furthermore, a new mindset means that technology is used to do something different, as compared to an old mindset, where well-known things are done but with more, or other, technology. With a new mindset, something different, like a multimodal text, could be created with the use of technology. This is however not a practice that has been widely adopted in education. Currently, formal learning in school is dominated by the reading and writing of traditional texts, whereas informal learning, often connected to activities outside of school, encourages young people to develop other skills and competences in meaning-making which involves producing and consuming multimodal and multimedia texts (Kress, 2010). This change cannot be ignored by formal education, instead the challenges that this presents for formal education need to be addressed, in order to provide adequate education for the future.

In research on literacy, a need for new theoretical ways to discuss reading and writing skills has been discerned (e.g. Alexander & Fox, 2004; Skaftun, 2010). We argue, among others (e.g. Adami, 2015; Lemke, 2006), that this discussion needs to be widened and applied to communicative skills as a whole in trying to answer questions such as: What does it mean to be able to express meaning in contemporary societies? What competences will be needed in the future to be an active citizen? Which subjects in school should be responsible for developing the students’ wider literacy competences? How can new media technology be incorporated in curricula and syllabi? These are all questions that are important to native language teachers in particular.

Jewitt (2006) points out that by using ‘literacy’, the old, hierarchical way of grading modes is maintained. Contemporary communication in diverse modes and media, among a diversity of rapidly changing social and cultural contexts, need new and flexible competences and skills, defined as multiliteracies (e.g. Kalantzis & Cope, 2012). To talk about ‘multiliteracies’ instead of just ‘literacy’ builds on a socio cultural and critical understanding of ‘literacy’, but also adds the multimodal and multimodal aspects of communication. Literacy practices in education are affected by the institutional context and therefore the activities that teachers and students engage in need to be related to this setting. Whether meaning expressed in a variety of modes, will be recognized as learning depends upon, and is affected by, rules in the educational setting, such as curricula, but also by traditions in certain communities, such as language teachers, and established ways of working within the subject (Selander & Kress, 2010). In order to understand, and perhaps explain, the literacy practices that teachers and students engage in in the language-classroom, they have to be related to these, often invisible, aspects and how they affect both the practice as such as and what is recognized as learning. This will in the following analysis be done with the use of analytical tools from Cultural Historical Activity Theory (hereafter CHAT) as conceptualized by Engeström (e.g. 1998).

---

1 The concept of prosumer can be used to talk about the combined role of both producer and consumer (Toffler, 1980).
2. Theoretical Framework – Multimodality and Activity Theory

Multimodal theory formation, developed since the 1990s, is undergoing a dynamic discussion concerning its possibilities, limitations, concepts and research possibilities (Jewitt 2014a; Jewitt 2014b). We consider it an apt perspective for discussions about the meaning of education and teaching in the media landscape of contemporary societies.

From a multimodal perspective, meaning-making is possible in different modes and media in a non-hierarchic, ecological way (Barton, 2007; Kress, 2010). All modes (in Kalantzis & Cope, 2012: audio, visual, gestural, tactile, spatial, verbal spoken, verbal written) have full meaning potential which means that a meaning-offering can convey ideational, interpersonal and textual meaning\(^2\) in its own right. This non-hierarchic approach to understand meaning-making can be used for discussions of how and to what extent different modes are used and recognized as learning in formal education both in, for example, teachers’ actual planning and in assessment.

In a socio-cultural perspective, learning is social, situated and mediated (Säljö, 2005). With a multimodal approach, the mediation applies to the meaning potential of all modes, i.e. meaning is made, distributed, interpreted and remade through many communicational resources (cf. Jewitt, 2008, Kress & Van Leeuwen, 2001).

In adapting a terminology for speaking about meaning-making through various and equal modes, design is a useful concept and a way to talk about meaning-making and learning simultaneously. Whilst meaning is actively designed by the individual, patterns and conventions are also inherited and affect the design process. In this view, teachers are regarded as designers of learning processes and environments and students’ as designers of their learning (cf. New London Group, 1996; Selander & Kress, 2010). Conceptualizing education and learning as a process of designing meaning-making differs from views on learning as a process of transferring knowledge (cf. Säljö, 2010).

Cope and Kalantzis (2010), Jewitt (2006), Kress (2003; 2010) and Kress & van Leuween (2001) all use multimodal design as an overall concept. Designing is the active process which is used to conceptualize meaning-making at different levels: in work with assignments (available design) students create (design) new meaning (redesign) (Kress & van Leeuwen, 2001). Meaning-offering, as mentioned above, is a concept that can be used regardless of mode or medium. Meaning-offering, as well as meaning-making, are related to the design process as shown in Table 1 below.

<table>
<thead>
<tr>
<th>Meaning-offering</th>
<th>Meaning-making</th>
<th>Representation/ new meaning-offering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available design</td>
<td>Designing</td>
<td>Redesign/ new available design</td>
</tr>
</tbody>
</table>

Talking about reading, creating and discussing texts as ‘design’, brings the whole design process within the focus of interest. It also means that reading is seen as one of several possible ways to make meaning (cf. Jewitt 2006). By viewing all modes as having full meaning potential, the multimodal approach questions the primacy of the written or spoken word in society in general, and in education in particular. Since multimodal texts are becoming increasingly important and common in the communicational landscape of today, this challenges the conception and evaluation of literacy in education and what is recognized as learning. From the theoretical standpoint of CHAT, activities are conceptualized as activity systems containing different components. The components at the local level of the activity system are the persons engaging in the activity, the tools that are used and the object, or outcome, of the activity. Engeström (1998) calls the subject, object and mediating tools “the tip of the iceberg” as they are “visible instrumental actions of teachers and students” (ibid., p. 79). The components at the systemic level are community, rules and division of labor (Engeström, 1987). These less visible components contain the structure of school systems, whereas the local level relates to content and methods of teaching (Engeström, 1998). Changes in these components lead to tensions and

\(^2\) Drawing on Systemic Functional Grammar (Halliday, 1985), the meaning potential in a multimodal perspective is applicable in all modes. Ideational, interpersonal and textual meaning refers to different functions of the meaning-offering. The ideational function expresses our experiences of the outer world. The interpersonal function deals with our relations to others and the textual function refers to how information is organized.
contradictions between the components and in and between activity systems and these tensions may, in turn, lead to change.

The tools that we use today for expressing meaning are often some kind of technological, digital device. With these devices, messages can be conveyed in a number of modes that were not possible when we mainly used pen and paper. This means that the tools and the object of activity systems have been altered and this affects the other components. When considering what is recognized as learning, the systemic components are important to consider as they relate to the historical aspects of classroom activities. In order for the outcomes of the activities to be recognized as learning, teachers, as a community, will have to recognize these outcomes as meaningful and part of the content of the subject. Moreover, these outcomes need to be related to curricula and syllabi in order for teachers to be able to assess what students do. Previous research has shown that the division of labor in the classroom is affected by the introduction of personal digital tools in classrooms (e.g. Tallvid, 2014), since these tools give students access to an abundance of information which may challenge the teachers authority and expand the object of the activity. Traditional roles where the teacher is the expert is thereby challenged and since students may use tools that are not familiar to the teacher to accomplish the object of the activity, the teacher’s ability to both support and constrain the actions of the students’ may diminish.

3. Multimodal Meaning-making – Example and Analysis

Assessment is an important factor of how learning is perceived and established (e.g. Drotner & Erstad, 2014). As shown in research (e.g. Godhe, 2014) assessment practices are crucial in determining what is recognized as learning. Other important aspects are the teaching and the assignments (Selander & Kress, 2010). In a multimodal perspective, assignments can encourage students to use multimodal meaning-making in their design process, and the extent to which they do so may depend on how the assignment is expressed and prepared through teaching.

The following example outlines the local level in the activity system and how students create meaning consisting of several modes with digital tools. A class in the second year of upper secondary school in Sweden is working with poetry in the subject of Swedish. The students are to examine what has been considered as poetry through history; who has been writing, what, how and for whom. This forms a part of the available design they have access to when they are given the assignment to create a poem of their own. In this particular class, the teacher had the aim of teaching from a multimodal perspective, trying to encourage the use of different modes for communication. This is noticeable in the assignment, which is also an important aspect in enabling for the recognition of the use of other modes than the verbal in the making of poetry. The students were given the assignment:

"Create poetry! What do you want to express? Can you express it in different ways?"

Another way (and, to our experience, a more common way) would be to give an assignment like;

"Write your own poem and illustrate your poem with a picture".

In the given assignment, there are no preference of mode ("create" instead of "write") and no definition of how to use different modes. O Instead, there is an opening towards the students own choice since direct questions are used in the assignment. This is, in our opinion, one aspect of the assignment which renders possible the recognition of multimodal meaning-making as learning. Below is an example of how a student chose to present the assignment.

In picture 1 the modes used are visual (color, size, background, vector) and verbal writing but in presenting the poem in class also verbal speech and audio (beginning with silence followed by a recorded voice that whispered the lyrics in a cold staccato voice). Moreover, the classmates were involved since the silence in the beginning resulted in students moving towards the screen to be able to read.
Based on the example, and our knowledge of the circumstances, we argue that the teacher tried to make it clear in the lectures what a multimodal approach to communication implies. Although this was new to the students, they show in their solutions that they have tried to use several modes and specific semiotic resources in creating their poetic meaning-offerings.

It is not unusual for teachers to encourage the use of different modes and digital media for student work but, as research show, it is more unusual to recognize it as learning both in the assignment and in assessment (Godhe, 2014; Oldham, 2005; Selander & Kress, 2010). In the example above, the multimodal perspective encourages questions about how the semiotic resources of the various modes are used and perceived by the students and accordingly the presentation in class involved the students explaining their choices of modes, the resources used and their intentions of making meaning and how the multimodal approach affected the meaning potential. The students’ reflections over the design process, was then used as a base for the assessment and grading. The teaching of poetry in this example is an example of how multimodal meaning-making can be recognized through teaching, the assignment and assessment.

To be able to understand how to use different modes, and the various resources connected to them requires competences that go beyond the competences needed for verbal meaning-making as writing, or oral meaning-making. The important question to ask is what this means for the competences needed as a language teacher?

3.1 Analysis; Levels of Meaning-making

In the design process meaning-making can be referred to at different levels, as shown in figure 1 below. On a general level, meaning-making concerns understanding meaning offerings in a similar way, regardless of mode or media. On the general level, the understanding deals with questions of sender, purpose of message and transmission, which concern overall questions about the ideational, interpersonal and textual meanings that can be perceived when examining the meaning offering on a general level (cf. Kress 2008). The general level is necessary, but today no longer enough. If meaning-making is considered and taught drawing on the idea that all modes are equal and have full meaning potential, there is also a need for an understanding of how the semiotic resources in the various modes work, which point to the specific level of meaning-making. For a language teacher, the specific meaning-making of verbal oral and writing are competences that are well established within the subject and which are known and practiced by both teachers and students. For example, in order to examine the specific level of verbal meaning-making in poems, the focus in the teaching is on voice, figurative
language, intertextual connections, meter and rhyme. Examining this specific level in a meaning offering promotes understanding of how the specific resources are used to create meaning.

Figure 1: meaning-making on different levels. From Tønnessen (2011).

To fully understand the meaning-offerings exemplified in the poems above the teacher and students need to understand the specific level, not only for the verbal mode but also for the audio, visual and spatial modes. In the example above, this would involve the need for discussions about, the size of the writing, the placement of the writing in the left hand corner, the color of the writing and background, and the silence followed by the digital voice and, additionally, how the different modes work together.

3.2 Analysis; Tensions and Contradictions

In the example above, it is evident that using digital devices to communicate, changes the nature of the message since digital communication facilitates multimodal communication. This means that the activity of reading and writing, in educational settings as well as in other settings, is affected (cf. Lemke 2006). In an educational context, these changes need to be related to factors at the systemic level in order to understand, and potentially explain, how and why tensions and contradictions occur when attempting to change practices (cf. Engeström, 1998).

Classroom practices are governed by rules such as the curricula. Syllabi and curricula are written based on conceptions of what different subjects contain and what students should learn in these subjects. Subject traditions affect what both teachers and students perceive as the content of the subject, which in turn influence how and to what extent digital tools are incorporated in the teaching (Selwyn, 1999). The content of the particular course in Swedish discussed in this article, for example, the content is specified in 11 statements. Only of these concerns multimodal meaning-making: “Configuration of texts for various purposes and media, including various multimedia texts where, for example, written text, pictures and sound interacts” (Skolverket, 2011).

Studies in Norway show that the expansion of texts to include multimodality, is not a notion that is self-evident or established by teachers in general (Matre et al., 2011). Hjukse (2010) points out the difficulties in assessing texts containing several modes and stresses the need for teachers to address these issues. In the Swedish curricula, recent changes has meant that the concept of a widened notion of texts has been removed and replaced by a more implicit understanding of texts as incorporating all modes. This unclear definition opens up for interpretations and, seen in the light of the Norwegian studies, is not beneficial for teachers’ assessment of multimodality. In the course syllabus referred to above, one out of five assessment criteria can be interpreted as pointing to multimodal meaning-making: “the student can […] configure various genres of text graphically and can […] motivate choice of illustration, disposition and over all outline” (Skolverket, 2011).
This leads us to the question posed earlier, what competences are needed by teachers today? Is it possible for teachers to exclusively focus on traditional meaning offerings like written texts, thereby ignoring a large proportion of the meaning-offerings that the students come into contact with on a daily basis? There are several possible ways to answer these questions and the answers have consequences for how contemporary meaning-making is dealt with within education. In our view, if subjects are to go multimodal all the way, there is a need for developing knowledge and competences about the specific semiotic recourses connected to other modes than the verbal. Remaining where most subjects are at present, teaching, assignments and assessment take into account general meaning-making for all modes, but only deal with the specific in relation to the verbal modes. This is going multimodal half way. There is also the possibility of only engaging in meaning-making that is verbal writing or oral on both general and specific levels. Then the multimodal meaning-making in subjects go no way. We argue that language subjects need to take into consideration various modes and media in teaching communication for today and the future. This means multimodality in subjects needs to go all the way and calls for further changes in curricula.

4. Conclusion and Discussion

In order to tackle what is sometimes referred to as a literacy deficit in younger generations, our suggestion is to widen subjects to fully embrace and recognize meaning-making regardless of mode and media. The purpose of this article is to discuss the challenges that face language education due to the rapid changes in the communicational language which has led to abundance of multimodal texts in contemporary societies. As stated by Cope et al. (2011:84), to use words alone is simply not enough to express meaning in contemporary societies. Based on the presented example, as well as previous research within the field, we will conclude by arguing for subjects in general, and language subjects in particular, to fully embrace and recognize multimodal meaning-making as important for students to understand and reflect upon, both as consumers and producers in the contemporary communicational landscape.

For this to happen, we argue that meaning-making in education needs to be based on a non-hierarchical and inclusive view on modes and media in order to create a readiness and flexibility for the demands of a rapidly-changing society, now and in the future. Taking this stance means that we have to deal with challenging changes in how subjects are conceptualized. Moreover, this has implications for teacher education and the professional development of practicing teachers since these educational efforts need to include and practically work with a widened concept of meaning-making within all subjects. Language subjects need to play an important role in developing students’ wider literacy competences, but multimodal meaning-making needs to be considered and evaluated in all subjects.

Elmfeldt and Erixon (2004) state that while writing is considered by both students and teachers to belong to language subjects, modes such as images and sound, are regarded as adhering to other more practical or esthetical subjects. Hobbs (2006) claims that incorporating media technology has been more difficult in language subjects than in other subjects and refers to this as a type of ‘digital divide’ to the detriment of language teachers. To change conceptions of what language subjects include, changes are needed in curricula, but also in the competences required by language teachers. While national Swedish curricula for primary and secondary school have been changed to strengthen students’ digital competence, there is currently no similar initiative to change teacher education. We argue that that there is therefore an apparent risk that the conception of what the subject of Swedish contains will continue to be based on verbal meaning making within teacher education. This will make the process of changing the conceptualization, and widening the content, of the subject amongst language teachers in general slower and more difficult. A major issue when discussing language subjects in Sweden appears to be students’ lack of interest and reluctance to read. However, considering reading in a broader sense, we nowadays consume and produce a considerable amount of texts on a daily basis and on top of that storytelling is abundant, if including the consumption of multimedia productions. By embracing multimodal and multimedial meaning-making and incorporating it as valuable learning in language subjects, students and teachers could access and assess students’ complete literacy competences, instead of focusing on some practices, while largely ignoring others. Recognizing the ways in which students consume and produce meaning would probably mean that students also find language subjects more interesting to engage in. By comparing and juxtaposing different kinds of meaning-making it would be
possible to discuss what can be gained and what may be lost by conveying meaning in a certain way. How does an instructional film-clip compare to a written instruction? What are the advantages and disadvantages of using different modes? What is it that it is possible to convey in written text and what messages benefits from using images and sounds? What is literature today? Are, for example, films and TV-series included in the concept of literature?

Raising these kinds of questions in classrooms and in teacher education would also make it possible to discuss and actually teach about different kinds of reading. As suggested by Kress (2010), different kinds of reading skills are required for different activities. Kress takes the example of game-players and argues that when playing a game, reading is required for obtaining information at a particular point in time. This kind of reading differs from the kind of reading that is generally in focus in schools. In education reading predominantly refers to the reading of longer texts where the reader is required to concentrate their attention on the text during extended periods of time (ibid.). According to Kress, this kind of reading can no longer define what reading is; instead it needs to be taught as a special task. Instead of regarding this kind of reading as the definition of what reading is, teachers and students need to explore what different kinds of reading mean, as well as when, and why, they are needed. Raising students, and teachers, awareness of the purpose of different kinds of texts and the reading of them is essential in order to develop the students’ ability to create meaning and become active citizens in the society of today and tomorrow.

The importance of taking a critical stance towards information in general has always been a salient part of literacy, but due to the tremendous accumulation of accessible information through digital devices, this competence is becoming increasingly important. In order for students to be able to take a critical stance in relation to information, whichever way it is mediated, the way the modes are assembled and how they affect each other and the message that is conveyed, needs to be discussed with students. As shown in the example in this article, such discussions can serve as the basis for assessment and evaluation of student productions. Furthermore, having these discussions about texts is a way to work with critical literacy so that students become critically aware of how meaning is made and are able to be critical towards messages in all shapes and forms.

When students are given access to an abundance of information through the use of different digital devices, the division of labor in the classroom is affected (e.g. Tallvid, 2014). Established roles where the teachers are able to support and constrain the students work may be challenged, as well as the teachers’ position as an expert in the classroom. Acknowledging these changes and working with the possibilities of having several experts in the classroom, we argue, may also be part of the solution. Due to the rapid development of different digital media, programs, applications and so on, it is impossible for any single teacher to keep updated with everything. However, applying the concept of distributed knowledge, as suggested by for example Lankshear and Knobel (2008), means that the teacher’s and the students’ combined knowledge needs to be taken into account in the classroom. The teacher may not know which applications to use but students may know and be able to teach each other and the teacher. This does not mean handing over the responsibility to the students, since the teacher needs to organize the activities in the classroom so that they are directed towards established goals and criteria that the students work to obtain. Similarly, the distributed knowledge amongst colleagues, at the local workplace as well as in networked communities of teachers, could also be considered and consciously worked with so that teachers work together to solve common dilemmas and to learn from each other.

Recent surveys of the digitalization in Swedish schools show that the access to digital devices in general is high, but the use of these devices is restricted in many subjects (Skolverket, 2016). Teachers consider themselves to be in need of professional development focusing on the pedagogical use of digital tools. We argue that it is of importance for professional development to focus on the general contributions that the digitalization of education potentially conveys and avoid focusing on ever-changing aspects, such as applications. Moreover, it is necessary to relate the use of digital devices to the subject in question in order to establish and discuss the particular ways that digitalization affect different subjects. Potentially positive, as well as negative, ways in which the content of the subject may be affected, needs to be related to in order to come to a workable solution of the pedagogical role that digitalization plays in educational settings. Furthermore, changes in the communicational landscape and how these may affect the subject, needs to be considered. This includes the concept of multiliteracies where meaning-making in different modes and media are evaluated and embraced in the teaching.
To sum up, we argue for a need to change the conceptualization of subjects in general, and language subjects in particular, so that they become multimodal all the way. In order to accomplish this, changes are needed in the rules that govern teaching, such as curricula and syllabi. Moreover, changes in how the community of teachers of languages regards the subject are needed. This involves practicing teachers, as well as those who are currently studying to become teachers, and therefore affects both teacher education and professional development of teachers. As part of the solution, changes in division of labor, both in the classroom and in the staff room, needs to be taken into account and worked with so that the combined knowledge of all parties involved is utilized. If language subjects do not seriously consider and evaluate contemporary meaning-making, they run the risk of becoming ‘encapsulated’ in the game of schooling (Engeström, 1991; Resnick, 1987). This means that activities carried out in the language subjects have little, or nothing, to do with the meaning-making that students engage in outside of school, since the use of digital technologies are added onto institutionalized practices instead of captivating the full potential that the changes in mediational means carry for meaning-making practices.

References


853
Effectiveness of a Learning Design Combining Summary-speaking Self-study Using Mobile Application with Paired Reflection on Learners’ Speaking Process

Kae NAKAYA* & Masao MUROTA

*Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Japan
bInstitute for Liberal Arts, Tokyo Institute of Technology, Japan

*knakaya@mr.hum.titech.ac.jp

Abstract: The authors earlier developed the Mobile Applications for Scaffolded autonomous summary speaking Task (MAST) for supporting learners’ self-study of English speaking. Although MAST enables learners to conduct summary-speaking tasks without teachers’ support and improves their oral fluency, it cannot offer enough feedback to enable learners to test whether the listener can comprehend their oral output, which is also necessary for learners to acquire second language. Therefore, we proposed a learning design that combines self-study using MAST and reflection activities in pair work in an English class. This research’s objective was to investigate the effectiveness of the learning design for learners’ oral performance and to clarify the role of self-study and reflection activities for development of learners’ speaking skills. After analysis, we concluded that the learning design was effective for improving learners’ oral performances. We also discussed the relation between the oral performances and learners’ cognitive process in our learning design.

Keywords: Computer assisted language learning, oral task, summary speaking, pair work, reflection, oral performance

1. Introduction

Because globalization continues to accelerate, people whose native language is not English face the important issue of improving their English skills. However, self-study especially practice English-speaking is difficult for English as a Foreign Language (EFL) learners without a teacher’s support.

To address this problem, the present authors developed Mobile Applications for Scaffolded autonomous summary speaking Task (MAST) (Nakaya and Murota, 2016a). In this system, learners speak an English newspaper article’s summary in English through self-study. The feature of the system is to offer scaffolding practice for speaking its summary. By conducting the practice, learners could understand the main points of the article and necessary words of its summary. We found that MAST enables learners to practice speaking English on their own and to improve their oral fluency.

Despite the effectiveness, learners needed more help to reflect on their speech. Using MAST, learners could check their recorded summary-speaking and compare it with a sample summary as a reflection activity. In the activity, the learners tended to reflect on their speech in terms of only fluency. Swain (2005) stated the importance of a peer’s feedback for second language (L2) acquisition. In order for learners to test whether a listener can comprehend their oral output and to seek better expression, feedback from other people seemed to be needed.

Therefore, in a previous study, we proposed a learning design that combined self-study using MAST and pair work in an English class (Nakaya and Murota, 2016b). With this design, learners practiced English summary speaking through self-study at home using MAST, and after that, they told the summary to a peer and next, reflected on it with the peer, using a worksheet designed for an English class. The worksheet facilitated the peer in clarifying and discussing improvements of learners’ summary-speaking step by step.
Although we found that the learning design above increased learners’ motivation for speaking English outside the classroom (Nakaya and Murota, 2016b), it remains unclear why the combined learning design is effective. Consequently, the objective of this research is to clarify the roles of self-study and subsequent pair-reflection activities by evaluating the learning design’s effectiveness for learners’ oral performance in detail, targeting Japanese undergraduate and graduate students.

Figure 1. Speech Production Model Based on Levelt (1989).

2. Background

2.1 MAST’s Effectiveness on the Speaking Process

In designing a method for supporting learners’ speaking skills, considering the influence of oral tasks’ features on learners’ cognitive process is important. Skehan (2009) attempted to explicate the relation among learners’ oral performances, features of oral tasks, and learners’ speaking processes based on Levelt’s model (1989). But before considering a MAST design based on Skehan’s suggestion, we first explain Levelt’s speech production model.

Levelt’s model (Figure 1) illustrates how, when speaking, people process information in three main stages: Generating preverbal messages in Conceptualizer (Figure 1-a), retrieving lexical information from Lexicon (Figure 1-b), and building syntactic structure in Formulator (Figure 1-c). First, Conceptualizer generates a preverbal message and sends it to Formulator. Second, Formulator retrieves the necessary words from Lexicon. Third, Formulator processes the information grammatically. Fourth, Formulator encodes the structure into a phonetic plan. Fifth, Articulator converts the phonetic plan into audible sound.

According to Skehan (2009), if the information that learners process in Conceptualizer is concrete and easy, they can pay more attention to process in Formulator. In addition, readiness for using necessary lexical items enables learners to retrieve them smoothly from Lexicon. Therefore, when learners’ cognitive load for generating a preverbal message and retrieving words decreases, they can allocate cognitive resources to grammatical encoding.

Scaffolding practice in MAST was developed according to Levelt’s model and Skehan’s discussion. MAST offers scaffolding that enables learners to understand a text’s main points and what kinds of words they can use to summarize it aloud. That is, learners do not have to pay much attention to generating a preverbal message (Figure 1-a) and retrieving the necessary words (Figure 1-b); thus they can focus on grammatical encoding (Figure 1-c). By repeating focused grammatical encoding, automaticity of processing linguistic form increases, in turn resulting in improved fluency (Nakaya and Murota, 2016a).
2.2 The Role of Reflection in a Learner’s Speech with a Peer

Swain’s (2005) output hypothesis indicated peers’ importance in L2 learning. If a learner has a partner when speaking English, the learner can test through their peer’s feedback whether her or his knowledge construction of L2 is correct. This process provides the learner opportunities to modify her or his output, resulting in L2 acquisition.

If the partner is a native speaker, the learner can obtain immediate feedback through interaction (e.g., recast; the native speaker correctly modifies what a learner has said). However, in classrooms where English is not a native language, most students are non-native speakers, and thus for a peer to provide immediate and appropriate feedback is often difficult.

Therefore, we designed a reflection activity on a learner’s summary speech with a peer in a classroom. After a learner speaks a summary in English, the learner and peer, through discussion based on a worksheet, gradually clarify the following points; (1) what the peer could understand during the speech, (2) what the peer could not understand, and (3) improvements to the learner’s speech, for example, other words, phrases or additional information to help the peer better understand the summary’s content. Following this procedure, the pair can consider together what and how the learner needs to improve in the speech so that the learner obtains valuable feedback in spite of pair work between non-native learners.

2.3 Literature Review of CALL Researches

In terms of supporting learners in improving oral performance, computer-assisted language learning (CALL) researchers evaluated effectiveness of technologies or learning designs that combine self-study using technologies outside the classroom and activities in the classroom.

As for effectiveness of technologies, many researchers clarified how and why technologies contribute to improvements of learners’ speaking skills. For example, spectrograms or pitch contours of a learner and a native speaker using automatic speech recognition could work as a visual feedback and improve learners’ pronunciation (e.g. Hardison, 2004; Olson, 2014). By offering such visual feedback, learners could recognize the gap between a native speaker’s pronunciation and the learner’s own pronunciation (Olson, 2014). Regarding synchronous computer-mediated-communication (SCMC), it can create learning environment similar to face-to-face communication. Previous researches have evaluated its effectiveness such as increasing amount of learners’ output (e.g. Abrams, 2003) and negotiation for meanings (e.g. Yanguas, 2010). Other researches have explored cognitive mechanism when using SCMC and influence on learners’ oral performance (e.g. Payne and Ross, 2005).

Another studies which focused on the learning designs have tended to demonstrate the effectiveness on learners’ perception or on their overall performance. Students conducted some learning activities using technologies such as watching English videos, and worked on oral tasks in pairs or groups in class. The combined learning design could ameliorate not only learners’ performance (e.g. Hung, 2015) but also their perception (e.g. Cheng et al., 2010).

In contrast, the final goal of our research is to clarify the role of self-study using technologies and the role of classroom activities, and how technologies support and maximize effectiveness of face-to-face learning activities. To achieve the goal, we investigated effectiveness of the combined learning design on learners’ oral performance based on Levelt’s model (Levelt, 1989).

3. Learning Design

In our learning design, learners first do self-study and practice English summary-speaking using MAST; then they conduct pair work in English class, as described below.

3.1 Learning Procedure using MAST by Self-study

Figure 2 shows screens and the learning procedure for MAST. A learner reads an English newspaper article (Figure 2-a) and summarizes it aloud (Figure 2-e). After reading the article and before the summary-speaking task, the learner conducts three scaffolding practices (Figure 2 b-d). In a week, the
Learner completes the five tasks (Figure 2 a-e) for one article and can repeat all the tasks as many times as she or he wants.

Details of MAST procedure are as follows: First, a learner reads an English newspaper article. During this activity, the learner can record some words (maximum of five) from the article by tapping on them (a). Words recorded in this step are shown on the screen during summary-speaking tasks (c - e), so the learner can refer to the words for the oral summary. Second, MAST offers Short question and answer practice (b). MAST vocalizes a question related to the article’s summary points, and the learner immediately voices the answer. After that, MAST voices a sample answer and then offers the next question. The learner repeats the pseudo-interaction five or six times. Third, the learner conducts two kinds of scaffolded summarizing tasks (c and d). During the learner’s summary speaking, MAST screen shows the previously recorded vocabulary list, all the sentences, and the picture in (c), the vocabulary list and the picture in (d). Finally, the learner summarizes aloud referring only to the vocabulary list (e).

For using MAST, we designed two kinds of scaffolding to decrease learners’ cognitive load for summarizing aloud. One scaffold is Short question and answer practice (b). By listening to the question, answering it, and listening to the next comment, learners can test their comprehension of the article, clarify its main points, and obtain hints for modifying the summary. Moreover, learners can conduct this practice as a pseudo-interactive conversation with a virtual tutor. The screen shows a female picture, and her face changes according to the dialog. Learners must immediately answer the tutor’s question. This procedure possibly decreases learners’ language anxiety and increases their motivation. The other kind of scaffolding is Fading summary speaking tasks (c and d). The article’s information on the screen fades step by step, so learners can try more difficult summary speaking in small steps, which leads to the main summary-speaking (e).

![Figure 2. The Screen and Procedure of MAST.](image)

3.2 Procedure of Pair Work in an English Class Using a Worksheet

Figure 3 is the worksheet designed for pair work. The picture (Figure 3-a) and the description (Figure 3 b-e) were written by one of the learners and a peer during an experiment described in section 4.

The detailed procedure using the worksheet is as follows. First, learners conduct a speaking task. A learner speaks a summary, which she or he practiced with MAST at home, to a peer. During the summary, the peer-listener draws a picture to express what the peer understands (a). Second, the pair reflects on the summary. They discuss what the peer-listener could and could not understand based on the picture. Third, the learner writes what the peer-listener could not understand (b) and detailed scripts that she or he spoke actually (c). Fourth, they think of other words or phrases that would help the peer-listener better comprehend the summary, and note the improved scripts (d). If the pair gets other feedback from the other pair or the teacher, they note it in (e).

The worksheet has two significant features. First, the picture drawn by the peer-listener helps the pair visualize the speaker’s learning achievement. By observing the picture, the speaker can confirm how much of the summary the listener could understand. This might provide objective feedback to the learner and lead to discussion on effective improvement of the summary. Second, the worksheet facilitates the pair’s reflection on the summary speaking and thinking step by step of detailed improvements.
4. Outline of the Experiment

4.1 Objective

This experiment’s objective was to investigate the learning design’s effectiveness on learners’ oral performance. We implemented pre- and post-tests with speaking tasks, transcribed the speaking tests, and evaluated them, using measures for oral performance. We describe the measures in section 4.3 below.

To clarify the effectiveness of a combination of self-study using MAST and the reflection activity during classroom pair work, we compared results of the learning design in this research with results of the previous learning design that offered neither scaffolding in self-study, nor the reflection activity with a peer (Nakaya and Murota, 2015). In the previous learning design, learners practiced summary speaking without scaffolding practice, using a mobile application at home, and then they explained the summary in English to a peer, but reflected on the summary independently. Therefore, we could utilize the previous learning design’s data as a control group. We attempted to clarify its effectiveness by observing not only test scores, but also learners’ output during pair work.

4.2 Schedule

We conducted an experiment for four weeks (29th October to 26th November, 2015). On the first day, we explained the experiment to learners and lent everyone a tablet PC (Nexus 7) on which MAST was installed. Using MAST at home for six days, learners practiced English summary speaking; on the seventh day, they conducted pair work in an English class. Learners were 25 Japanese first-year undergraduate students majoring in computer science. Two learners did not participate in the pre-test, and the other two learners had trouble in the pre- or post-test. Therefore, we excluded their data.
In June 2015, the previous learning design’s experiment was conducted in a different class of the same university, using the same framework. Six learners did not use the previous MAST at all, and one learner had trouble in the pre-test. Therefore, we excluded their data from analysis. For detailed information, please see Nakaya and Murota (2015).

4.3 Data Collection

As for learners’ oral performance, many researchers have measured three items: Complexity (Structural and Lexical), Fluency (Speech rate, Repair fluency, and Silence), and Accuracy (e.g., Tavakoli and Skehan, 2005; Mehnert, 1998). Based on these studies, we transcribed learners’ speeches in pre- and post-tests and calculated the following scores.

For complexity, we evaluated (1) Structural complexity measured by counting non-repeated words per the Analysis of Speech Unit (AS-unit) (Foster et. al, 2000), and (2) Lexical complexity measured by counting the ratio of unique words. Higher Structural complexity and Lexical complexity show amelioration in a learner’s oral complexity.

For fluency, we evaluated (1) Speech rate measured by counting the number of non-repeated words per minute; (2) Repair fluency measured by counting repetitions of exact words, syllables, or phrases, corrections, and partial repeats, and (3) Silence measured by counting silences that last 0.4 seconds or more. Higher Speech rate, and lower Repair fluency and Silence mean improvements of learners’ oral fluency.

For accuracy, we evaluated the ratio of error-free AS-units. A higher score of accuracy means that a learner can speak English with less errors.

When observing learners’ reflections, we counted how many noted detailed improvement (other words, phrases, or added information) on the worksheet. Figure 3 displays a typical example describing “other phrases.” The learner spoke “…approved a new reform for the framework of program”, and the learner and a peer thought up the other phrase: “…change the kinds of sports at the Tokyo Olympic.” They tried to use another easier phrase to help the peer’s better comprehension of the summary. If a learner wrote more than one kind of improvement, we counted both.

5. Results and Discussion

We analyzed all data on learners’ oral performances by mixed two-way repeated measures ANOVA using two variables. The between-subjects factor was “learning methods,” which had two levels (the current learning design and the previous learning design), and the within-subjects factor was “test term,” which had two levels (pre- and post-test). Table 1 shows the results.

5.1 Results in Complexity

Figure 4 displays results for Structural complexity and Lexical complexity.

For Structural complexity, marginal differences were revealed in the “learning method” factor ($F(1,34)=3.288, p<0.1$), “test term” factor ($F(1,34)=3.436, p<0.1$), and interaction ($F(1,34) = 3.942, p < 0.1$). Therefore, we conducted tests of simple main effects. Results for the “test term” factor in the experimental group showed significant differences ($F(1,34) = 7.370, p < .05$), and Structural complexity in the post-test was higher than in the pre-test, as shown in Figure 4-a. Results for “learning method” in the pre-test showed significant differences ($F(1,68) = 6.916, p<.05$) and the control group’s scores were higher than the experimental group’s scores.

Lexical complexity showed no significant differences, as shown in Figure 4-b.

5.2 Results in Fluency Scores

For Repair fluency, significant differences were shown on the “learning method” factor ($F(1,34) = 4.430, p < .05$). The control group’s scores were higher than the experimental group’s scores. Figure 5-a shows the results.
For Speech rate, significant differences were shown on the “learning method” factor ($F(1,34) = 4.454, p < .05$) and the “test term” factor ($F(1,34) = 25.764, p < .001$). These results show that both groups’ scores improved, but the experimental group’s scores were higher than the control group’s scores. Figure 5-b shows the results.

For Silence, significant differences were shown on the “test term” factor ($F(1,34) = 1211.925, p < .001$) and marginal differences were shown on interaction ($F(1,34) = 57.407, p < .1$). Therefore, we conducted tests of simple main effects. Results for the “test term” factor in the experimental group and the control group showed significant differences ($F(1,34) = 44.871, p < .001$ for the experimental group; $F(1,34) = 18.524, p < .001$ for the control group), that is, both groups’ scores improved. Figure 5-c shows the results.

It is worth noting that improved Speech rate and Silence scores of both groups showed that summary-speaking by self-study and telling the summary to a peer, which were conducted by both groups, were effective for these two scores. The improvements were shown on Figure 5-b and c.

In addition, note that statistically significant differences between learning methods in Speech rate and Repair fluency just showed the differences between the groups’ initial ability. The results showed that experimental group scores in all terms were superior to the control group scores. In other words, students in the experimental group could speak English more fluently from the beginning. The result was due to the fact that students of the experimental group and control group attended different classes.

Table 1: Mixed Two-way Repeated Measures ANOVA.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>DOF</th>
<th>Mean Square</th>
<th>$F$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>Between learning methods</td>
<td>11.340</td>
<td>1</td>
<td>11.34</td>
<td>3.288</td>
<td>0.079+</td>
</tr>
<tr>
<td></td>
<td>Among test terms</td>
<td>6.059</td>
<td>1</td>
<td>6.059</td>
<td>3.436</td>
<td>0.073+</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>6.952</td>
<td>1</td>
<td>6.952</td>
<td>3.942</td>
<td>0.055+</td>
</tr>
<tr>
<td>Lexical</td>
<td>Between learning methods</td>
<td>260.521</td>
<td>1</td>
<td>260.521</td>
<td>2.516</td>
<td>0.122</td>
</tr>
<tr>
<td>complexity</td>
<td>Among test terms</td>
<td>96.755</td>
<td>1</td>
<td>96.755</td>
<td>2.137</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>6.209</td>
<td>1</td>
<td>6.209</td>
<td>0.137</td>
<td>0.713</td>
</tr>
<tr>
<td>Speech rate</td>
<td>Between learning methods</td>
<td>1285.808</td>
<td>1</td>
<td>1285.808</td>
<td>4.454</td>
<td>0.0423*</td>
</tr>
<tr>
<td></td>
<td>Among test terms</td>
<td>1030.779</td>
<td>1</td>
<td>1030.779</td>
<td>25.764</td>
<td>0.000****</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>75.034</td>
<td>1</td>
<td>75.034</td>
<td>1.875</td>
<td>0.18</td>
</tr>
<tr>
<td>Repair</td>
<td>Between learning methods</td>
<td>227.390</td>
<td>1</td>
<td>227.390</td>
<td>4.430</td>
<td>0.0428*</td>
</tr>
<tr>
<td>fluency</td>
<td>Among test terms</td>
<td>3.709</td>
<td>1</td>
<td>3.709</td>
<td>0.267</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>4.690</td>
<td>1</td>
<td>4.690</td>
<td>0.337</td>
<td>0.565</td>
</tr>
<tr>
<td>Silence</td>
<td>Between learning methods</td>
<td>10.350</td>
<td>1</td>
<td>10.350</td>
<td>0.091</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>Among test terms</td>
<td>1211.925</td>
<td>1</td>
<td>1211.925</td>
<td>60.528</td>
<td>0.000****</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>57.407</td>
<td>1</td>
<td>57.407</td>
<td>2.867</td>
<td>0.0996+</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Between learning methods</td>
<td>0.000547</td>
<td>1</td>
<td>0.000547</td>
<td>0.017</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>Among test terms</td>
<td>0.000263</td>
<td>1</td>
<td>0.000263</td>
<td>0.011</td>
<td>0.918</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>0.0264</td>
<td>1</td>
<td>0.0264</td>
<td>1.078</td>
<td>0.307</td>
</tr>
</tbody>
</table>

5.3 Results in Accuracy

For Accuracy, there were no significant differences.
Figure 4. Results for Complexity.

Figure 5. Results for Fluency.

Table 2: The number of reflection on the worksheet (N = 21).

<table>
<thead>
<tr>
<th>Reflection</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
</tr>
<tr>
<td>Other words, phrases</td>
<td>8</td>
</tr>
<tr>
<td>Added more information</td>
<td>10</td>
</tr>
<tr>
<td>The peer could understand the speech completely</td>
<td>0</td>
</tr>
<tr>
<td>Grammar correction</td>
<td>1</td>
</tr>
<tr>
<td>Improve body language or intonation</td>
<td>2</td>
</tr>
<tr>
<td>Nothing</td>
<td>2</td>
</tr>
</tbody>
</table>

5.4 Results of Descriptions on Reflection Worksheets

Table 2 shows learners’ reflections for each week. In reflections, many learners could write detailed improvements using other words or phrases, or adding more information. The result might show that learners could consider more options for expressing what they wanted to say so that the peer could understand more easily.

5.5 Discussion

The learning design could improve Structural complexity but not Repair fluency. The reasons of the results are discussed in the following paragraphs.

In Structural complexity, learners who conducted self-study with scaffolding of MAST and pair reflection could use more words in one AS-Unit, as Figure 4-a illustrates. In other words, the experimental group learners tended to add more words or phrases to explain what they wanted to say. We provide typical examples in pre- and post-tests and explain details in the next paragraph.
In the pre-test, Learner A of the experimental group spoke in very short sentences, “…many people, many people have a little time at lunch time. So, we can eat Ramen for short time. So it is good for lunch to eat.” In the post-test, Learner A spoke in this way, “…I like to, I like to take a train and go to another, another place, especially most place Tohoku and Hokkaido.” Learner A explained her or his hobby by adding one more phrase. Another example shows that Learner B spoke in this way in the pre-test, “…everyone practiced hard for this. And I practiced hard, too, for it,” but this way in the post-test, “The thing that I, I feel the most enjoy is, hmm, entry the contest of the brass band.” Learners A and B used more than one verb and added an adverb phrase, an adjective clause, or a noun clause, so that the number of words in one AS-Unit increased.

This effectiveness might show that the reflection activity could facilitate learners in gaining strategies of how to think up and use approximate phrases when they do not retrieve appropriate words to express what they want to say. We can explain it by referring to Kolb’s (1994) experiential learning model. Learners in the experimental group spoke the summary in English, and then they reflected on their speech with their peers. During the reflection activity, many learners discussed better expression that enabled their peers to understand their speech, as seen in Table 2. According to Kolb’s experiential learning model, when learners have concrete experience and reflect on the experience from a different perspective, they can construct more abstract knowledge. In other words, reflection with a peer stretches the retrieval of words from Lexicon.

MAST scaffolding for summary speaking also has an important role in gaining such strategies. Mackey (1999) observed that in language learning, feedback’s effectiveness depends on learners’ developmental readiness. Feedback from peers or instructors does not work if the learner does not have appropriate readiness. In this research, learners who obtained MAST scaffolding during self-study might have activated knowledge about vocabulary needed for summary speaking because the scaffolding facilitated learners’ use of recommended words in the summary through Short question and answer practice. This might have helped develop readiness, which led to greater effectiveness at stretching learners’ knowledge of how to retrieve words from Lexicon.

The unchanged Repair fluency of the experimental group, which differs from the previous result using only MAST (Nakaya and Murota, 2016a), might have been caused by the higher Structural complexity scores. As mentioned previously, experimental group learners might have learned how to access Lexicon from another perspective. Thus, they might have tried to search and retrieve better lexical items or phrases for additional explanation to express what they wanted to say, and sometimes repeat the same words when searching the words. In the post-test, Learner A spoke in this way, “…and go to another, another place, especially most place Tohoku and Hokkaido,” which we showed the previous paragraph. During repeating the word "another", the learner might try to search a phrase to add more information, which is "especially most place Tohoku and Hokkaido." This kind of utterance resulted in keeping Repair fluency scores, as shown in Figure 5-a. In addition, although there were no significant differences between pre- and post-test scores of the control group, Repair fluency scores of the control group seemed to improve, which might be due to repeating the summary-speaking task.

6. Conclusion and Future Works

In this research, we investigated the effectiveness of the learning design that combines scaffolded summary-speaking self-study using MAST and pair reflection in an English class. After analysis, we concluded that the learning design was effective for improving Structural complexity, Speech rate, and Silence scores. The current learning design compares favorably with the previous design, which was effective only for increasing Speech rate and decreasing Silence and which offered only summary-speaking tasks and speaking to a peer.

From these results, reflection on a learner’s speech with a peer might enable the learner to stretch knowledge of how to access and choose vocabulary words because, during reflection activities, learners received feedback from a different perspective for expressing what they wanted to say. In addition, scaffolding practices in self-study play an important role in gaining strategies because scaffolding might activate related knowledge for expression as a pre-task.

In general, future studies need greater clarifying investigation on the effectiveness of reflection activities and scaffolding practices separately. Moreover, we can seek out broader computerized possibilities for assisting language learning, with ICT playing some of the roles mentioned previously.
For example, technologies such as voice recognition and corpus studies might offer functions to develop readiness for interaction with other people by judging learners' achievement of oral tasks and recommending more appropriate phrases during self-study using ICT. Such practice activities might make EFL learners' experiences in face-to-face communication more effective.

Acknowledgements

We would like to thank Prof. Masatoshi Tamura and Prof. Takehiko Tanioka of Institute for Liberal Arts, Tokyo Institute of Technology for their help in conducting the experiments.

References

Discussion Course Model Using Online Educational Resources to Enhance EFL Learners’ Motivation and Critical Thinking

Yuichi ONO* & Ai NAKAJIMA

*Center for Education of Global Communication (CEGLOC), University of Tsukuba, Japan

School of Engineering, Tokyo University, Japan

*ono.yuichi.ga@u.tsukuba.ac.jp

Abstract: This paper concerns the implementation and utilization of online educational resources (OERs) in a course model that aims to enhance Japanese EFL learners’ motivation and critical thinking. The course model incorporates components of OERs, flipped teaching, OER learning support systems, and digital storytelling. Following a four-week project, the course model was found to have influenced students’ motivation and critical thinking.

Keywords: Online educational resource (OER), flipped classroom, OER learning support system, digital storytelling, critical thinking, motivation

1. Introduction

There is general agreement among foreign-language researchers and instructors that course models should meet the requirements of “vigorously changing societies” in the twenty-first century (Chan, 2010; Kong, 2014; Gut, 2011). To achieve this goal, foreign-language courses must develop information literacy and critical thinking skills, along with communicative proficiency, in the target foreign language. The growth of digital culture in the twenty-first century drives both instructors and students to use online educational resources (OERs). Despite the challenges of controlling difficulty levels, instructors have shown increasing interest in new, authentic Creative Commons–licensed video content such as TED Talks, YouTube, and so on. The study reported in this paper focused on a flipped discussion course for Japanese university students using a supporting system to help students work with game-based vocabulary learning through OERs, equipped with automatic quiz generation and learning portfolios. On the basis of a four-week pilot experiment, we conclude that the course model—still in progress—can enhance learners’ motivation and critical thinking.

2. Background

2.1 Online Educational Resources (OERs) as Teaching Materials

Following the long history of the “open movement” for educational resources (Nakajima & Ono, 2014; Ono & Nakajima, 2014), massive open online courses (MOOCs) and online educational resources (OERs) have been adopted worldwide (Allen & Seaman, 2014). Among the top 100 universities, 81 initiated their own MOOCs in 2015 (Shigeta, Sakai, Tsuji, Inaba, & Hiraoka, 2016), suggesting that MOOCs have become standardized educational resources. On the other hand, OER use is not quite as widespread. Allen and Seaman (2014) note that in the United States in 2014 only half of higher education institutions were working to actively use OERs. There are several reasons why OERs have not been extensively utilized thus far.

Three aspects should be considered in order to capture the characteristics of OERs for use in foreign-language teaching: (i) authenticity, (ii) diversity, and (iii) multimodality. Regarding (i), most will agree that the materials should be authentic. However, topics that are too highly contextualized and
specific might reduce students’ motivation to learn something new. Similarly, students can learn the various complexities of linguistic forms or the diversity of English as indicated by (ii). However, it might be difficult for beginners or those who are not interested in the topic to obtain complex linguistic knowledge. Regarding (iii), instructors are willing to employ videos available online. For example, TED Talks have become popular online resources for discussion or debate tasks in Japanese secondary-school foreign-language courses. TED videos are direct examples of presentation models used as teaching materials and visual aids for learning authentic English and other linguistic aspects. Ono, Nakajima and Ishihara (2017) argues for the merits of diversity of OER in terms of students’ self-study based on their interest by observing how students choose the TED videos.

2.2 Digital Literacy and Critical Thinking

Information literacy and critical thinking are two important skills to possess in the twenty-first century (Kong, 2014; Gut, 2011). Processes specific to information literacy include gathering, synthesizing, analyzing, interpreting, and evaluating information (Kong, 2007; Price, Becker, Clark, & Collins, 2011). Collecting OERs on a single topic leads students to perform these processes to create and construct new knowledge.

Critical thinking is the ability to think reflectively and make skillful judgments about what information is reliable and what actions should be taken (Ennis, 2002). This paper largely adopts Hirayama and Kusumi’s (2004) formulation of critical thinking in relation to Japanese students. According to their quantitative research, critical thinking comprises four subcategories: (i) awareness of logical thinking, (ii) inquisitiveness, (iii) objectiveness, and (iv) evidence-based judgement. In addition, Hirayama and Kusumi (2004) claim that belief bias negatively affects learners’ ability to draw logical conclusions. However, more “inquisitive” students are able to escape belief bias.

2.3 Learning Support System

The learning environment in this study utilized the learning management and support system YouTutors (Nakajima & Ono, 2014; Ono, Nakajima & Ishihara, 2017). The system starts with an automatic quiz generator (AQG), which allows instructors to save time developing online quiz modules for vocabulary. The instructor simply has to copy the URL of a TED Talk and paste it into the box in the interface. After receiving feedback from users and students, we made some important modifications: (i) a more user-friendly interface, (ii) learning management system (LMS) functions such as class management and portfolios, (iii) more gamification factors, and (iv) learning-behavior visualization. The system is able to connect to an online dictionary. Ono and Nakajima (2017) quantitatively and qualitatively examined YouTutors’ effect on learners’ motivation. Our revised system interface is shown in Figure 1.

The game mode adopted here is a “fill-in-the-blank” mode, which we can also call a “typing” mode. In this game, users listen to a video and type the words they hear. The first version had two other distinct modes; however, we will skip these because they are not relevant to our research goal. Figure 2 shows a screenshot of the typing mode.
2.4 Course Design

The course design is composed of the following components: (i) flipped classroom, (ii) jigsaw-based discussion, and (iii) digital storytelling (DST).

2.5 Flipped Classroom

The flipped classroom model is a pedagogical approach to blended learning in which the traditional model of classroom lecture followed by homework is reversed and often supplemented by instructional videos (Garrison & Vaughan, 2008; Khan, 2012; Tucker, 2012). In our course model, instead of conducting traditional face-to-face lectures in the classroom, instructors select ready-made OERs, or educational videos on the Web (e.g., Kahn Academy, TED Talks, iTunes U, YouTubeEDU). In the
actual lecture, the instructor acts as a facilitator for students, who engage in various problem-solving activities that require them to apply the knowledge they acquired by completing their homework. This approach facilitates a more efficient use of class time in which the classroom is more “active” and “communicative.”

2.6 Jigsaw-Based Discussion

Jigsaw-based discussion is a cooperative learning strategy. The content of the lesson is subdivided into different pieces of information. Then, it is given to groups of students who explain their respective pieces to each other, thus completing the jigsaw puzzle (Aronson & Patnoe, 1997). Figure 3 shows our teaching model in which the subtopics are given as questions about information from the text.

In the first step, students form “expert” groups in which they focus on one subtopic or question, and then research and discuss it. The students become “experts” on their assigned subtopic or question. Then, the students from all of the “expert” groups form a new group to teach their peers based on their findings and discussions. Eventually, all members of the groups will have learned from each expert group discussion, obtaining mutual benefits.

2.7 Digital Storytelling (DST)

The main effects of DST are illustrated in various contexts. Previous studies of DST suggest that creating digital stories can promote cognitive development, self-authoring, and identity construction (Davis, 2004; Sadik, 2008). Other merits of DST include enhancing technical skills, engaging students, and sharpening critical thinking skills (Castañeda, 2013; Sadik, 2008). For teachers, they can easily assess students’ progress toward learning goals since their work is constantly recorded and reflected in every process of learning (Ono, 2014). L2 writing and multiple speech-draft recording tasks involving DST enhance learners’ awareness of linguistic skills, especially writing and speaking skills (Castañeda & Rodriguez-Gonzále, 2011). Ono (2014) noted that DST reduces speaking anxiety for less confident EFL learners since they can avoid standing in front of people face-to-face by using presentation software. Ono (2014) also suggested that for high-level learners willing to speak in public, DST enhances project-based-learning (PBL) skills such as computer use and searching (computer skills); collecting, sorting, and analyzing information (information literacy); and problem-solving and critical thinking skills (academic thinking).

2.8 Course Model

Figure 4 shows the course model integrating the three abovementioned components. The flipped discussion class was conducted for three weeks (weeks 1-3). Each week, students were required to study TED Talks with YouTutors. In this project, the topic was nuclear power. The first week focused
on TED Talk presentations on the dangers of nuclear power. Then, during the second week, students studied TED Talks arguing for the positive aspects of nuclear power. During week three, the students watched any clips of TED Talks that supported their own opinions. Finally, in the fourth week, after gathering, synthesizing, analyzing, interpreting, and evaluating the information obtained through flipped discussion, students created videos presenting their opinions. During the discussion sessions in weeks 1-3, several foreign students joined the groups to facilitate the discussion.

![Figure 4. Course Flow.](image)

![Figure 5. TED Talk Videos for Week 1 and Week 2.](image)
3. Study

3.1 Research Question

This study aimed to evaluate the implementation of a four-week course in terms of students’ motivation and critical thinking. Specifically, we attempted to answer the following research questions through pretest- and posttest-designed mixed research:

RQ1: Does the course model involving OER, flipped discussion, and DST have an effect on students’ motivation?
RQ2: Does the course model involving OER, flipped discussion, and DST have an effect on students’ critical thinking?
RQ3: How do students evaluate the use of OER, YouTutors, jigsaw discussion, and DST?

3.2 Participants

Forty first-year university students in Japan participated in the study (male: 16; female: 24). Since the class was created based on the results of TOEFL placement tests, the proficiency levels among the participants were controlled.

3.3 Procedure

Regarding RQ1 and RQ2, questionnaires were used before and after the project. For RQ3, a postsurvey was conducted using open-ended questions. Regarding learners’ motivation, we adopted the ARCS model (Keller, 2010), which holds that learner motivation can be analyzed in terms of four independent aspects: attention, relevance, confidence, and satisfaction. The Instructional Materials Motivational Survey (IMMS) was used as the questionnaire scale (Huang & Yoo, 2010; Keller, 2010). It consists of 36 items on a five-point Likert scale, divided into four basic components. For RQ2, we used the Critical Thinking Disposition Scale for Japanese Students (Hirayama & Kusumi, 2004). This scale has 33 questions in the following four categories: awareness of logical thinking, inquisitiveness, objectiveness, and evidence-based judgment. Average scores were compared using a t-test.

3.4 Results

For the motivational survey, the following results were obtained, as shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Attention</td>
<td>2.99</td>
<td>0.32</td>
<td>3.17</td>
</tr>
<tr>
<td>Relevance</td>
<td>2.65</td>
<td>0.30</td>
<td>2.82</td>
</tr>
<tr>
<td>Confidence</td>
<td>2.73</td>
<td>0.45</td>
<td>3.06</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3.27</td>
<td>0.45</td>
<td>4.23</td>
</tr>
</tbody>
</table>

Note: ***p < 0.001; **p < 0.01.

It can be observed in Table 1 that the students improved their awareness of motivation to some degree. Among the four motivational components, “satisfaction” significantly improved with an average score higher than 4.0.

Next, we will look at the results of the critical thinking survey (Table 2).
Table 2: Means, standard deviations, and t-tests for critical thinking dispositions.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Awareness of logical thinking</td>
<td>2.86</td>
<td>0.38</td>
<td>3.33</td>
</tr>
<tr>
<td>Inquisitiveness</td>
<td>3.74</td>
<td>0.57</td>
<td>4.45</td>
</tr>
<tr>
<td>Objectiveness</td>
<td>3.11</td>
<td>0.45</td>
<td>3.66</td>
</tr>
<tr>
<td>Evidence-based judgement</td>
<td>3.27</td>
<td>0.78</td>
<td>3.61</td>
</tr>
</tbody>
</table>

Note: ***p < 0.001; **p < 0.01.

Table 2 shows improvement in the values for all four categories. Among them, “inquisitiveness” showed a significantly high score of 4.45.

Regarding RQ3, participants were asked to write freely about each component of the course. Most participants wrote positive remarks about each component (e.g., “The task was useful”). Table 3 shows common feedback for each component.

Table 3: Common feedback comments.

<table>
<thead>
<tr>
<th>Component</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ted Talks</td>
<td>Very difficult to understand/ Japanese translation was helpful/ The presenter speaks so fast/ 15-minute movie is too long to work on …</td>
</tr>
<tr>
<td>YouTutors</td>
<td>Very interesting/ Dictionary function is very useful/ Visualization is amazing/ I couldn’t get connected to the YouTutors from home/ Sometimes it does not run smoothly…</td>
</tr>
<tr>
<td>Jigsaw Discussion</td>
<td>It was fun to listen to other students’ opinions/ It was very difficult for me to prepare what to speak in the discussion/ Too short to complete discussion about all the questions…</td>
</tr>
<tr>
<td>DST</td>
<td>Time-consuming/ It was exciting/ I practiced speaking again and again/ Fun to watch other students’ movies…</td>
</tr>
</tbody>
</table>

4. Discussion

The results for RQ1 and RQ2 suggest that the proposed course model influenced learners’ motivation and critical thinking. A closer look at Table 1 shows that the score for “satisfaction” was very high while “attention” was over the median. This indicates that good implications are drawn from the ARCS model since these two components are crucial for continuing the cycle in the motivational process.

Regarding critical thinking, we observed that the score for “inquisitiveness” was very high. According to Hirayama and Kusumi (2004), belief bias hinders students from drawing proper logical conclusions. However, more “inquisitive” students can overcome belief bias. The data shown in Table 2 indicate positive consequences for critical thinking. Needless to say, to make students “inquisitive,” instructors need to provide proper topics and questions for discussion.

Importantly, the comments in Table 3 provide a lot of hints for improving the model in the near future. Specifically, the system needs to be modified so students can operate it smoothly and comfortably, since some students found it very difficult to follow the difficult Ted Talks video and it took much longer time to complete the vocabulary task. In addition, students generally had difficulty listening to the scripts in the video material, even though they liked choosing the video. These things seem to suggest that the system needs to be modified so that the gap between the task difficulty and students’ proficiency level can be minimized by introducing and incorporating student’s learning process data.
5. Conclusion

This study investigated implementing and utilizing online educational resources (OERs) in a course model aiming to enhance Japanese EFL learners’ motivation and critical thinking. The results were affirmative, but the open-ended question data indicated that the model requires modification. This study has limitations in that the sample size was small, and the experiment covered only four weeks. In addition, we need to further consider the nature of the critical thinking skills that might be appearing in students’ performances. In conclusion, the course model has room for improvement, and the best-utilized course model should be considered in the future.

References


Abstract: Too often educational literature constructs the use of digital technology as a communication-mediating device rather than as a tool with qualities that can be utilised for assisting learning. To counter this trend, it is necessary for educators to approach innovation in computer-assisted learning as an opportunity to question how we understand the challenges that the learning process presents to the students and how they can be addressed with the help of digital technology. This paper argues that transformative learning must begin with, and revolve around, a student examining his or her way of relating to the world; this is very different from apprenticeship models, popularised in the 19th century, that begin with a specific task or practice and train students in its performance. This is a conceptual paper. It describes reading tools especially designed to support first and second language students in developing literacy skills involved in the production of convincing and critically examined texts of all kinds, including public presentations or even poems. The paper demonstrates the relevance of these tools to transformative pedagogies, compares them with other literacy support generally available in schools, and illustrates how they can be integrated into a learning context. The emphasis of emotions draws on evidence from neuroscience, where both comprehension and learning are increasingly presented as emotional processes. It follows that the traditional thinking of texts as products of a cognitive process alone needs re-thinking to account for emotions as the source of intentions which inform both the process of text construction and its interpretation. It is therefore not the vocabulary that students struggle to comprehend or manipulate, but the ways in which emotions are turned into a text to evoke a desired effect in the interlocutors. Furthermore, an explicit attention to emotions allows students to approach literacy and language-learning as a whole person, expanding students’ emotional understanding of themselves and how they perceive others.

Keywords: Emotions, student wellbeing, text analysis, resource-based learning

1. Introduction

The task of learning to read and write, whether in one’s own or second language, has always been perceived as a challenge in educational literature, with “wars” being waged about the right way to approach text (Luke, 2000). The goals of a language and literacy classroom are said to be many. Typically, the literature speaks about students being able to engage their own cultural knowledge to “make language and concepts more meaningful to students” (Windle & Miller, 2012, p. 320-321), teachers learning about the students, discussions, sharing experiences, and writing activities that “focus on prior knowledge”, helping students to “recognise structural conventions and patterns” of texts, making comparisons “with meaning systems in other texts and cultural discourses” with students relating texts to their own experiences and feelings. To support critical analysis of texts, students are expected to “analyse issues and problems arising from a text”, draw inferences and explore generalisations (Windle & Miller, 2012, p. 321). Overall, the aim is for these literacy-learning experiences to result in students “developing skills for understanding how power is exercised through discourse” (Windle & Miller, 2012, p. 321). A formal analysis of the way in which texts are constructed is presumed to achieve this purpose (Windle & Miller, 2012, p. 321; Luke & Freebody, 1999).

The preferred way to teach is frequently referred to as a method of scaffolding, a gradual build-up of the skills necessary for independent writing. Among the scaffolding strategies are “teacher modelling and deconstructing of relevant text types”, “joint” teacher-student text construction (Windle & Miller, 2012, p. 321; Gibbons, 2009). The aim is for students to be trained by experts in order to engage “with the key ideas and concepts” like experts do (Windle & Miller, 2012, p. 321; Gibbons,
Students are also expected to “transform what they have learned into a different form for use in a new context or for a different audience” (Windle & Miller, 2012, p. 321; Cummins, 2000, 2001). To this end, teachers help students “translate their understandings into concrete actions, identity[...] roles that students can play by way of an intervention into an issue or problem” (Windle & Miller, 2012, p. 321). Students are given opportunities to engage their identities “through discussion of social and moral issues and through writing” (Cummins, 2011).

Discussion between teachers and students is viewed as the cornerstone of the scaffolding pedagogy (Windle & Miller, 2012, p. 321; Gibbons, 2009; Hardman, Smith, & Wall, 2003). It is believed that discussions allow students to “talk through complex ideas and hold substantive conversations”, all to induce students’ involvement. As suggested by Gibbons, 2009 (as cited in Windle & Miller, 2012, p. 322) the inclusion of pair and group work are advised as best suitable for this purpose. When testing the effectiveness of the abovementioned strategies in an ESL (English as Second Language) context in a study that involved thirty-nine Victorian schools for refugee-background students in Australia, Windle & Miller (2012) found vocabulary to be the single component that all teachers recognised to be important if students are to progress in their language and literacy skills. Discussions and questioning were identified as the key routines for bridging the students’ understandings and those of their teachers. Some other popular strategies involved considering students’ existing knowledge of English and the language demands of the new work, varying tasks or resources to support weaker students’ understanding of concepts, challenging students with work that is slightly ahead of their ability and support them to successfully attempt it (Windle & Miller, 2012, p. 324).

Windle & Miller’s study (2012, p. 326) showed also a great preference for teacher-led activities. They concluded that in the absence of “written resources at an appropriate level for students” teachers may lack alternatives, “We can’t use standard textbooks with these kids. We have to develop diversified curriculum materials which make the material accessible. (Science teacher)” (Windle & Miller, 2012, p. 327). Windle & Miller offer no comments about the nature of the resources that would be of value to those refugee-background students. However, some questions need to be raised. For example, if scaffolding is considered to be more student-oriented, what principles for material development does the scaffolding method offer to support “teacher modelling and deconstructing of relevant text types and joint teacher-student construction” (Windle & Miller, 2012, p. 321)? Furthermore, at the core of any strategy of text modelling and joint teacher-student text construction there is a belief that the process opens room for the students to find answers to the questions they experience when grappling with a task at hand, but do they? Neither Windle and Miller (2012), or any other sources that they reference, provide any form of evidence to support this claim. This is problematic considering that scaffolding routines form the majority of strategies on which modern education relies. Scaffolding routines are part and parcel of a pupil-teacher method of apprenticeship, first formally introduced to education in the 19th century (Carr cited in Thomas, 2007, pp. 4-5). In other words, the rationale supporting the model may draw more on tradition than critically-informed alternatives.

2. Critical Literacy in Practice

In order to contextualise the abstract steps of the process outlined above, Luke (2004) offers an example. The example is to show how students can be helped to recognise structural conventions and patterns of texts, analyse issues and problems arising from a text, and draw inferences to demonstrate how power is exercised through discourse. His analysis uses passages from a geography textbook.

information technologies and media and, most recently texts of new workplaces” (e.g., Kress & Van Leeuwen, 1997 as cited in Luke, 2000, p. 5; New London Group, 1997 as cited in Luke, 2000, p. 5). The intellectual premises of these principles Luke (2000, p. 6) outlines in reference to specific scholars, including Voloshinov and Bakhtin, Foucault, Derrida, Bourdieu, and Freire. In Table 1, the left-hand column illustrates Luke’s analysis (2000, pp. 11-12). The right-hand column in Table 1 provides the researcher’s responses to his concept of critique using the frameworks of these scholars. Luke’s analysis examines an introduction to a textbook in geography. He organised the text of this introduction by lines and offered comments relating to those lines. The analysis is a response to a following hypothetical classroom task:

As a tutorial group, go through the following textbook extracts and discuss the particular discourses and ideologies that are at work. Here are some key questions to briefly ask of each: • Which/Whose version of events and the world is foregrounded? • Which other versions are excluded? Whose interests are served by this representation? • How (lexically, syntactically, etc.) does the text construct ‘reality’? • How does the text try to position you in relation to its messages?

Table 1: A critique of Luke’s (2000) model for critical literacy.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 We will start by discussing some general types of impacts on the physical environment.</td>
<td>According to Luke (p. 6), “Voloshinov and Bakhtin’s views [are] that instances of language use are … instances of ‘heteroglossia’ where differential ideologies, struggles over difference and unruly social relations come into play”. It would follow that to critique is to engage in that struggle. However, the tutorial task does not insert the students into any context where they would feel a struggle within themselves and, as a result, the need to respond. A critique requires a purpose and this purpose is born from a tension that may not necessarily be addressed by questions such as those in the tutorial task. In short, the task is not framed to invite an exploration of one’s place among the social relations in which one is embedded. Instead, its questions train students as to how to read texts, not how to read one’s own position in relation to texts. Arguably, understanding one’s position would have a transforming effect on the student. However, when a task does not begin with a student’s own engagement in the world, it is not this engagement that is investigated by the student but arbitrary relationships that students are now asked to see as significant.</td>
</tr>
<tr>
<td>2 We will then look at particular types of environments, such as coastal Luke: “In lines 1 &amp; 3: who is ‘we’?”</td>
<td>Luke (p. 6) argues that “Foucault’s view [is] that discourse takes on a life of its own, constructing peoples’ identities, realities, and social relations; that is, that we are produced by discourse as much as we are producers of discourse”. To follow Luke’s argument, if it is so, then there is value in an exploration of one’s place among those discourses considering that that we are their producers and reproducers, i.e. more than</td>
</tr>
<tr>
<td>3 Lands, alpine areas, arid lands and cultural sites. Luke: “Lines 3 &amp; 4: are all the items on the list equivalent, opposites, dominant? How do ‘cultural sites’ fit in with the other things on the list?”</td>
<td></td>
</tr>
<tr>
<td>4 The construction of a resort or a complex of resorts and facilities is the most obvious of tourist impacts. Luke: “Line 5 &amp; 6: ‘most obvious’ to whom? Note the nominalisation “construction”, who is doing the construction? Does the nominalization (a verb turned into a noun) hide the agent, the ‘doer’ of the action?”</td>
<td></td>
</tr>
<tr>
<td>5 Very significant changes in land use result. 7 Areas of natural vegetation might be cleared. 9 Existing land-uses such as agriculture might be displaced. Older 10 parts of cities might be demolished … Luke: “Lines 5-10: note use of passive: where is the agency (developer), Who is doing this?”</td>
<td></td>
</tr>
<tr>
<td>6 One very important fact to remember is that natural environments Luke: “Line 11: Who is supposed to ‘remember’”</td>
<td></td>
</tr>
</tbody>
</table>
12 (such as mangroves, rainforests and water catchments) are valuable
13 in their natural states.
**Luke**: Lines 12-13: Notice again the list: are these equivalent items?
14 They all play a part in the ecological processes from which we
**Luke**: “Line 14: Who is we?”
15 benefit.
**Luke**: “Line 15: Recognised by whom?”
16 It is now recognised that ultimately these ecological processes
17 provide economic benefits.

just reproducers. But this is not how Luke sees it. He says, “this translates into a classroom focus
on identifying the dominant cultural discourses ... in texts” (Luke, p. 6). In this belief, Luke loses
the individual as a producer and, therefore, the entire contexts of relationships that impact on his
or her selections.

Luke (p. 6) continues, “Derrida’s view [is] that texts cannot be the objects of definitive interpretations, but involve the play of inclusions and exclusions, presences and silences”. While
there may be no cannon to warrant a single interpretation, comprehension is also not rule-free. Understanding what students include, and why, can have a transformative effect on
them. But there is also no single “identity-bound” interpretation either (Calhoun, 1995). Reading, therefore, is not about linking interpretation with an identity (“focus on multiple possible ‘readings’ of texts”, Luke, p. 6). A more productive way is to support students
in their task of linking their interpretations with the gains that are afforded by manipulating
different forms of capital (Bourdieu, 1991).

Luke (p. 6) believes that “Bourdieu’s view [is] that language is one form of cultural capital with
variable exchange value in social fields of institutions and communities”. However, neither
the fields nor the values are self-contained. They are dynamic constructs, and Bourdieu’s
“thinking tools” (Bourdieu as cited in Jenkins, 1992, p. 67) were developed to capture these
dynamics from the perspectives that they offer to an analyst. This implies that the social relations,
Sources of power and authority, of the institutions (e.g., mass media, workplaces, corporations, governments, educational institutions) cannot be identified as Luke
suggests independently of the interests (forms of capital) that the analyst brings with him or her
into the context of analysis. The task of analysis is part of the interplay of these interests; this is
why an analysis of an arbitrarily selected text cannot result in a transformative impact as it
does not engage forms of capital that can only be identified when their power (i.e. the capacity to
elicit the desired impact) is challenged.

Luke (p. 6) also draws on Freire, who believed that “that literacy education can generate tools
and conditions for people to reposition themselves in relation to economies, cultures and
dominant ideologies”. However, this view cannot be taken out of content. It calls for a
pedagogy where critique is not conducted ‘for its own sake’, but to solve problems. This is not exactly how Luke understands it. He believes that a perspective or a culture is representative of a specific group. Luke (p. 6) encourages the production of texts “from a range of cultures and institutions”. In so doing, he makes it difficult to imagine how change happens, i.e. how an individual could change his or her mind, or “culture”, if they belonged to a specific group. Indeed, how could individuals even understand other cultures if they function within the boundaries of their own limitations? Swales (1993, p. 695), himself once a proponent of this model, criticises it for its inability to explicitly capture “forward momentum or the pursuit of novelty: new ways of doing business, new genres, new subject matter, new product, the creation of a new research space” and, thereby, the concept of agency.

While the questions identified by Luke in the left-hand column (Table 1) may appear to make sense to any person trained by this method, the analysis in the right-hand column shows that these questions, together with the tutorial task proposed by Luke, model an “act of compliance to school routine” (Kramsch, 2000, p. 149). What they fail to do is to inspire students “to imagine and effect active intervention in the situations that affect them” (Freadman, 1994, p. 21). However, to inspire, it would be necessary to abandon traditional school routines and to engage students in contexts which would provoke in them the need to respond to the conditions that affect them. Any type of support (including activities, materials or people) would then function as sources of perspectives from which they evaluate their own positions and the “profits” (Bronfenbrenner, 1979), or forms of capital (Bourdieu, 1991) that validate it. A systematic analysis therefore would not be about an assessment of what others say or do, but about identifying and evaluating texts that bear directly on the response of the students and, thereby, on the conditions that generated the need for this response.

This is a very different approach to critique, i.e. one which is both analytical and creative, and where comprehension or intervention are conceived as processes that help students to adopt an informed position. An informed position is constructed by purposefully seeking and linking what previously seemed irrelevant or incommensurable (Calhoun, 1995; Latour, 1999). This process of seeking and linking requires the construction of new questions and therefore new perspectives from which one’s analysis is both examined and, therefore, understood. Arguably, a critique can have a transforming impact when individuals better understand their own position in the contexts that affect them.

3. Reading for Emotion: Reconciling the Concepts of Learning and Student Agency

In order to provide exploratory and dynamic learning conditions that were argued for in the previous section, it is necessary for researchers to interrogate different conceptual frameworks for their potential to allow students to examine the beliefs that motivate their actions from the perspective of different gains, that these make possible, and their various combinations. Turning these frameworks into learning tools is another matter that needs consideration in relation to pedagogy. The tools presented and discussed in this section are based on the understanding that in order to act informed, students need access to resources and tools that help them review and analyse what people do, why, how and when, from the perspective of their own questions and the challenges that these explorations help them identify. Since we are never either inside or outside a “culture” (Freadman, 1994), explorations of this
kind make it possible for students to creatively engage the different patterns they discover across different genres and experiment with them to elicit the desired communicative impact (Lian & Norman, 2017; Lian, A-P., 2004). Technology can help with such compare-and-contrast activities because technology-enhanced learning environments can be designed to enable students to (a) explore what they know; (b) identify the limits of this knowledge, and (c) generate new forms of knowledge and new possibilities.

In this resource-based model of learning, the idea is not to create texts that are assumed to be appropriate for specific subjects and types of students. Instead, the aim is for researchers to design tools which can increase interactivity of learning environments by making use of random access capacities of digital technology, practically unlimited storage of digital information, and of research in a diversity of fields and areas that can stimulate creative and critical thinking about the kinds of difficulties that students may be facing when working with texts. Students of all ages can then access a variety of resources and approach them from the perspective of their own needs and challenges; there is no need to place arbitrary limits on what they can read, analyse or do. They can identify their own boundaries.

The tools described in this section build on the previous attempts which sought to provide students with access to a set of multimodal examples of ‘life situations’, catalogued in the form of a database, with all records categorised to offer relatively rich information relevant to these situations, as well as increase flexibility of retrieval of the records. The records in the database had an additional function. The idea was for students to interrogate the database, ask for precise examples of interactions and request a specific lesson program (amongst other things) based on those selections. For example, when preparing radiobroadcasts, students could ask the database for examples of questions from a radio interview to practise intonation patterns of questions. Figure 1 illustrates a record from this database.

Figure 1. The Audiovisual Database. (Lian & Lian, 1997)

One of the shortcomings of the information presented in the database presented in Figure 1 was its inability to offer students a way of thinking about text organisation that they already “know” innately and that may therefore prove to be more helpful than the details identified in those initial database records. Emotion is one such organising factor. According to neuroscience (Damasio & Immordino-Young, 2007), the brain structures that organise our social behavior are built on the emotional structures that are more ancient. In order to explore how texts communicate emotions, at least one other feature of texts had to be identified for the proposed approach to be of value to students. For this purpose, a generic structure of all texts was identified and proposed as the second element. Based on years of analysis (Lian, 2006), it was hypothesised that most well-developed texts shared a general
structure. In schools, typically, texts are presented as having “beginnings, middles, [and] ends” (Luke, 2000, p. 11). However, this is not a very helpful way to think of a text as one cannot create a text simply by being told to construct a “beginning”, or a “middle”. Instead, the following generic structure was proposed: Setting the Focus, Disturbance, Dialogue, Development, Resolution, and Morale. The same structure may be possible to be applied to posters and other non-word based texts.

When linking emotions with a generic text structure, a student may ask himself or herself: “What emotion is the author trying to evoke in the reader when setting the focus of the story?”, “How does this stage of the text make me feel?”, or “What devices is the author using to make me feel this way?”.

These questions can be followed by other questions which can encourage further explorations: “Why do you think you feel this way?”, “Where else have you seen a text starting in the same way?”, “Did you feel the same way then?”, “How else could the author express this emotion?”, etc. There is no end to the questions that can be asked in order to expand investigation. Structures like world knowledge, vocabulary, or different grammatical devices can also be discussed. But these are devices secondary as they serve to communicate emotions. Thus, rather than being taught a specific process of knowledge construction (e.g. “knowledge processes”, Cope & Kalantzis, 2009, p. 185), or even to focus on abstract concepts such as fields of social, cultural and economic power, students are taught to explore “what texts do” by approaching them as emotional constructs. They can compare and contrast the same and different genres of texts by examining emotions in relation to their place in the generic structure of a text. In the process, they can build increasingly informed hypotheses about texts and how they can be used, not to mention expanding their own emotional understanding of themselves and how they perceive others. The transformative impact of this kind of analysis is potentially enormous as the process challenges students to look deeply into themselves as producers and interpreters of texts about themselves and others; the analysis positions students as emotional beings in relation to other people, who too have emotions and respond emotionally to texts and contexts. This approach is also very different from the standard questions in classrooms of the kind like: “What does it mean?”.

Investigations of the kind described above are also helpful in the process of text construction. For example, it is way easier to identify an emotion for setting the focus of a poem than simply sitting down to write one. This is not to say that there is no learning curve involved in the learning to read for emotions. Studies show that children in school have a very poor emotional vocabulary, feel unable to name emotions and, consequently, address them in a constructive way (Joseph & Strain, n.d.). However, an exploration of how authors communicate emotions, or reflections on how differently one could respond in the same situation, they all enable students to develop greater awareness of communication as emotions, take a distance to their own emotions and those of others, explore them, imagine possibilities and re-frame contexts and likely impacts. The skills that can be learnt are beyond the scope of this article. When combined with other multisensory awareness-raising tools (Yang, Wannaruk & Lian, 2017), students are given tools that challenge, and assist the development of, skills that only recently are becoming recognised for their value to students’ personal and social lives.

4. Reading for Emotions and Technology

Windle and Miller’s (2012) study raised the problem of resources. However, as argued in the previous sections, these resources do not always need to be specially crafted for students from non-English language background. All students can benefit from imaginative solutions, text and non-text based. The reading for emotion method lends itself to a number of technological solutions, the obvious being a database with examples of texts of the same and different genre, analysed to exemplify the link between the generic structure of texts and emotions. The pilot database (Figure 2) of this kind was created by the researcher for the purpose of demonstrating to teachers how they can create a “Do-It-Yourself” systems, with the help of freely available databases, including Google Sites, to store examples of texts for students to explore. When linked to multimedia, including Youtube, directing the video to a specific place on the timeline, the database becomes a handy tool for students and teachers equally. It provides a quick access to a diversity of examples for students and teachers to investigate when needed. It is not only teachers that can create entries into the database. Texts analysed by students themselves can be entered by the students individually or as class, depending on students’ age and abilities.
5. Reading for Emotions and how it Works

The limited scope of this article does not allow for a detailed description of the experiences with the model. A brief example follows. When given a task to retell a text in class to classmates, students frequently copy texts from Wikipedia without their teachers’ knowledge. This makes texts look perfectly grammatical and yet, in terms of their communicative value, they are badly designed. Wikipedia tends to present texts in a factual manner, not always suitable for a presentation, for example, by an 11-year-old student to his or her peers. For example, when Wikipedia starts its description of the Pompeii tragedy with words to the effect: “On 24 August AD 79, Mount Vesuvius erupted in Pompeii”, it is likely that 11-year-old classmates would lose interest in the story right at the start. To assist them in developing their own stories about Pompeii while also using the Wikipedia text, students can be invited to explore the emotional effects of the Wikipedia text. For example, in relation to the “Setting the Focus” stage of the text, they can explore alternatives more suitable for their classroom peers. Students could have fun with this task and explore whether they want to surprise the audience (“You would never guess…”), unsettle (“Tragically …”), or intrigue (“2000 years ago …”). Matching emotions with relevant expressions is challenging and access to examples would certainly help in this respect.

Interestingly, preliminary experimental work shows that students tend to have a great difficulty identifying the moral of the stories that they analyse or seek to tell. In stories that draw on unfamiliar cultural frameworks, even the teachers find this task challenging. The “reading for emotion” approach also appears to make aesthetic aspects of texts visible. For example, an analysis of the story, “How the kangaroo got its pouch” (Cloudskipper Dreaming, 2011), showed the text to be very repetitive, reproducing the pattern of Dialogue and Development, not reaching the Resolution stage for quite a while. Only when reflecting on the moral of the story it became apparent that the repetitive structure was mirroring the enduring kindness of the mother kangaroo, whose ability to love was being tested by the “spirit”. Furthermore, other initial tests of the model indicate that the “reading for emotion” approach can prove helpful in oral presentations, helping students build their oracy as they try to consciously understand the emotions of the audience that they seek to affect. The approach also helps testing the coherency of these presentations as the moral of the stories that students communicate should align with the earlier stages of text production.

6. Reading for Emotions and its Limitations

As mentioned above, the study by Windle and Miller (2012) reported beliefs of teachers from different subjects, including science. It is likely that teachers may question the value of the “reading for emotion”
approach in science study. However, on reflection, how does a scientific text communicate satisfaction, inspiration, curiosity, or the feeling of clarity to the audience? These questions need to be explored and tested. The aim of this paper was to present the general idea of the approach. There is a need for further research in order to examine the possibilities of the proposed approach and its relevance in educational contexts.

7. Conclusion

It was the objective of this paper to show the relevance of emotions in the process of texts construction and text analysis. To make this point, two different approaches to text analysis were compared. The method presented by Luke (2000) and Windle and Miller (2012) was shown to present text analysis as independent of the specific purpose that motivated the need for analysis in the first place. It was argued that in Luke’s approach, students are constructed as entities caught in the three-dimensional space of assumptions about their identity, culture and economic factors, with text being conceived as means by which authors manipulate these relationships. The assumption underpinning this method is that students can be taught this process of manipulation as if the rules were independent of the individual contexts of the interlocutors and, instead, were governed by some more abstract, “social” principles. In this perspective, the individual is subsumed by the social, while the social space has no subjects. Text analysis is a reflection of one’s standpoint determined by a specific cultural, social, and economical group.

In contrast, the “reading for emotion” model approaches texts as emotional constructs, designed to evoke emotions in readers. In this context, text analysis revolves around an analysis of emotions and the devices that were, and were not, engaged by the authors. In the “reading for emotion” model, the position from which a student interprets a text is not determined by a specific social space they are claimed to occupy, but by the dynamics that they bring with them into the context of reading. It is therefore critical to acknowledge these dynamics and to recognise that “text comprehension” has less to do with a text at hand and more with the “conversations”, or purpose, from the perspective of which a specific text was engaged in the first place. Transformative learning was constructed as a reflection upon those conversations. This exploratory component of the “reading for emotion” method makes it inherently critical and creative, and brings the student back into the context of analysis, without denying the social basis of his or her experiences.

Some preliminary experiences with the model were mentioned. In its early stages of conception, the model showed to be of value in a number of contexts. Still, it is apparent that a database of texts analysed using the “reading for emotion” model could assist students, and others, with examples of texts analysis that are systematically developed using a wide range of texts and genres. The database could be enriched with other analytical tools, all to helps students reflect on the relationships that partake in text construction. The “reading for emotion” approach therefore does not attempt to model text analysis or text construction. Instead, it is a reflective tool proposed for students to explore the value of insights that they may gain when viewing texts as emotional devices. Research is currently underway in Australia investigating the relevance of the “reading for emotion” model to language and literacy skills of students and academics in Indonesia.

References

Damasio, A., & Immordino-Yang, H. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *International Mind, Brain, and Education Society, 1*, 3-10.


Automatic Classification of Teacher Feedback and Its Potential Applications for EFL Writing

Gary CHENG*a, Shu-Mei Gloria CHWOb, Julia CHENC, Dennis FOURNGc, Vincent LAMc, & Michael TOMc

aDepartment of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong
bDepartment of Applied English, HungKuang University, Taiwan
cEnglish Language Centre, The Hong Kong Polytechnic University, Hong Kong
*chengks@eduhk.hk

Abstract: This paper presents and discusses the initial data from a project that aims to develop a system for automatic tracking of student responses to teacher feedback in draft revision. One main purpose of the project is to design and implement a method for automatic classification of teacher feedback on students' draft essays in the EFL context. In this paper, we propose the automatic classification method and evaluate its performance in terms of accuracy. Our findings show that an accuracy of over 96% was achieved when classifying teacher feedback using the proposed method. They also show that the classification results could be analysed with other sets of data such as assessment grades to help teachers reflect on their use of feedback types and refine their feedback practice. This study can provide a basis for future research into automatic analysis of the impact of various feedback types on student revision.

Keywords: Automatic classification, teacher feedback, student essay, EFL writing

1. Background

Feedback is regarded as one of the most significant influences on student learning and achievement (Hattie & Timperly, 2007). In the English as Foreign Language (EFL) context, feedback has a great potential for improving the quality of student writing (Hyland, 2003; Cheng, 2017). A large body of research focused on the role corrective feedback (also known as grammar correction) plays during the process of writing. However, evidence from prior research was inconclusive as to the effectiveness of corrective feedback on student writing. Some studies (Truscott, 2004; Truscott, 2007) argued that this feedback type has little value for EFL development and it can even have an adverse effect on students’ ability to write accurately, while some others highlighted the importance of grammatical correctness in academic writing and suggested that it can benefit students’ writing (Ferris, 2004, Lee, 2008).

In addition to grammar correction, research was undertaken to identify what other characteristics of teacher feedback are important to substantial and successful revisions of student writing. It was found that longer and text-specific feedback (e.g. criticism on an issue) was more effective in encouraging student revisions than shorter and general feedback (e.g. positive comments) (Ferris, 1997). It was also found that the success of student revisions was associated more strongly with the types of problems identified by the feedback (e.g. incorrect lexical choice, lack of explanation and insufficient details) than the syntactic forms of the feedback (e.g. declarative, question and imperative) (Conrad & Goldstein, 1999).

However, the findings about teacher feedback were primarily drawn from a limited number of studies and student cases in English-speaking countries (Chiu & Savignon, 2006; Ferris, 2003). They may not be generalised to other populations where EFL students are living in their non-English speaking hometown such as Hong Kong and Taiwan. Besides this, the methodology of previous studies was based on manual classification of teacher feedback. It is not appropriate for analysing a large data set of written comments and may give rise to a problem of consistency in classification.

Given the limitations of prior research, this study sought to propose and evaluate an automatic approach to classify teacher feedback on student essays from an EFL writing course at a Hong Kong
university. It also aimed to explore the relationship between feedback types and assessments. Results of this study can offer insights into the effectiveness of using an automatic approach to identify different types of teacher feedback. They can also demonstrate the potential of using the automatic approach to examine the correlation between the number of feedback in different types and the grades given by teachers on student essays. With this initiative, teachers can be provided with some evidence on their use of feedback types on which feedback types they use are weakly linked with their assessments.

This paper is organised as follows. Section 2 proposes the classification framework for teacher feedback. Section 3 provides details of the automatic classification method for teacher feedback. Section 4 describes the research methodology of the study. Section 5 presents and discusses the results obtained from analysing the data collected during the study. Finally, conclusion and future work are given in Section 6.

2. Classification Framework for Teacher Feedback

To characterise the ways that teachers frame their feedback, Straub (1997) identified six categories of teacher feedback. The categories include (1) praise, (2) criticism, (3) imperative, (4) advice, (5) open question, and (6) closed question. Chen and Hamp-Lyons (1999) noted that Straub’s (1997) taxonomy of feedback categories was derived from the L1 (first language) context and it mainly focused on the syntactic forms of the feedback, so they added two more categories to fit into the EFL context: mechanics (i.e. the feedback that deals with grammar and punctuation) and ‘?’ (i.e. the feedback that conveys a meaning of ‘do not understand’). The current study adapts and extends the models of Straub (1997) and Chen and Hamp-Lyons (1999) by taking into account both text-specific features (e.g. content and language use) and syntactic structures (e.g. imperative and question forms) of teacher feedback on student essays. The proposed classification framework is shown in Table 1.

Table 1: Classification of feedback types.

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Praise</td>
<td>Positive comments, non-controlling</td>
<td>Well written</td>
</tr>
<tr>
<td>T2</td>
<td>Criticism</td>
<td>Negative comments or evaluations, authoritative</td>
<td>Confusing</td>
</tr>
<tr>
<td>T3</td>
<td>Imperative</td>
<td>Comments that tell the student writer to do or change something, usually starting with a verb in the imperative form</td>
<td>Be consistent</td>
</tr>
<tr>
<td>T4</td>
<td>Advice</td>
<td>Suggestive comments often in conditional mode</td>
<td>Maybe you could add some details here</td>
</tr>
<tr>
<td>T5</td>
<td>Closed question</td>
<td>Questions that either get a ‘yes’ or ‘no’ as answer, or else a simple one-word answer</td>
<td>Do you think you have given an adequate evaluation?</td>
</tr>
<tr>
<td>T6</td>
<td>Open question</td>
<td>Questions that require more than a ‘yes’ or ‘no’ answer, often starting with ‘what’, ‘where’, ‘why’, ‘who’, when’, and ‘how’</td>
<td>What does this mean?</td>
</tr>
<tr>
<td>T7</td>
<td>Content</td>
<td>Comments that often deal with the clarity and meaning of content, ideas and views</td>
<td>Some ideas need further elaboration</td>
</tr>
<tr>
<td>T8</td>
<td>Language use</td>
<td>Comments that often deal with the grammar, punctuation, spelling and word choice</td>
<td>Reword this</td>
</tr>
<tr>
<td>T9</td>
<td>Organisation</td>
<td>Comments that often deal with the organisation of ideas and linkage between sentences or paragraphs</td>
<td>In general, the ideas flow well</td>
</tr>
<tr>
<td>T10</td>
<td>Referencing and formatting</td>
<td>Comments that often deal with citations, quotations and references</td>
<td>Non-academic sources</td>
</tr>
</tbody>
</table>
3. Automatic Classification Method

The automatic classification method is based on matching teacher feedback against syntactic rules and semantic words extracted from a set of manually annotated data. In this study, a total of 3412 teachers’ written comments on 90 students’ draft essays were collected to build the data set. According to the proposed classification framework, each teacher comment was manually annotated by two researchers. Discrepancies between the researchers were resolved by discussion to reach consensus on the annotation standard.

The basic unit of annotation was a single sentence. Every sentence in a comment was marked up with two feedback types. One type was concerned about the form (T1 to T6) and the other was about the aspect (T7 to T10). Figure 1 shows a sample student text with annotated teacher comments.

On the syntactic side, every sentence was processed with a part-of-speech (POS) tagger to assign parts of speech to each word (Bird, Klein, & Loper, 2009). The most common sequences of POS were identified and extracted to form a set of distinctive syntactic rules for each feedback type. On the semantic side, a word-by-type matrix was constructed based on Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998). Each feedback type was characterised by a vector of word weights.

To classify a new comment, we first apply the POS tagger to it at the sentence level. Every tagged sentence will be matched against existing syntactic rules to determine its feedback types. If there are no matches, the sentence will be semantically transformed into a vector of word weights. The vector will subsequently be classified into a feedback type where their cosine similarity is the highest among other types.

4. Research Methodology

4.1 Study Context

This study is part of the research project namely “Towards automatic tracking of student responses to teacher feedback in draft revision”. It was undertaken at the English Language Centre of a Hong Kong university in the first semester of the academic year 2016/17. Participants of this study were students attending the Advanced English for University Studies (AEUS) course. AEUS was a 13-week, credit-bearing course that required students to research for, write, plan and revise an academic position argument essay, and to defend their views and engage with those of others clearly and logically in an mini oral defence. As part of the course requirement, students had to submit two academic position argument essays. The first was a 600-word draft, and the second was a polished, final essay on the same topic of 1200 words.

4.2 Participants

Ninety-two undergraduate students (30 males and 62 females) enrolled on AEUS gave their written consent to participate in this study. Their ages ranged from 17 to 21 years (M=18.15 and SD=0.94). They came from 6 class groups and 5 academic disciplines: Advertising Design (1 group), Accounting and Finance (1 group), Mental Health Nursing (1 group), Nursing (2 groups), and Physiotherapy (1...
Three English language instructors (IA, IB and IC) who taught the participating class groups were also involved in this study. Details of the participating class groups can be found in Table 2.

4.3 Data Collection and Analysis

Students in the AEUS course were required to submit a 600-word draft essay on a topic of their choice by drawing on academic sources such as peer-reviewed journal articles at Week 7. They received letter grades and written feedback on their draft before preparing and submitting their final essay by Week 11. The assessment criteria and weighting for the draft essay were: content (30%), organisation (20%), language (30%) and referencing (20%). Primary data of this study were the feedback on the draft and the assessment results. The feedback was categorised manually by researchers and automatically by a classification tool implemented in Python (Bird et al., 2009), respectively. Accuracy of the automatic classification method was measured by the proportion of machine classifications that agree with manual classifications. Five-fold cross-validation was performed, with four-fifth of the annotated feedback extracted as the training data and the remaining as the testing data. This procedure was repeated five times until all feedback in the original data set was tested once. The classification accuracy was calculated as the average over the five iterations. Additionally, letter grades in assessments were converted to numerical scores following university guidelines (A+ = 4.5, A = 4, B+ = 3.5, B = 3, C+ = 2.5, C = 2, D+ = 1.5, D = 1, F = 0). The results of classifications along with assessment scores were used to calculate the correlation between feedback types and assessments given by teachers.

5. Results and Discussion

5.1 Accuracy of the Automatic Classification Method

To contrast the actual and predicted classifications of teacher feedback and report the accuracy of the automatic classification method, a two-way contingency table known as a confusion matrix is shown in Table 3. In the confusion matrix, each row (or column) refers to the count of a feedback type identified by human (or machine). The numbers of correct classifications (i.e. the results of machine classification are identical to those of the manual classification) are represented by diagonal cells, while the numbers of mis-classifications are represented by off-diagonal cells.

A confusion matrix is a good way to illustrate the accuracy of the automatic classification method. As can be seen in Table 3, the proposed method could identify different types of teacher feedback with high degrees of accuracy ranging from 96% upwards. The results are very encouraging and suggest that it is feasible to classify teacher feedback in an automatic way.

5.2 Correlation between Feedback Types and Assessments

Pearson correlation coefficients were calculated between the number of feedback in different types and the assessment grades. Table 4 provides a summary of the correlation coefficients. It shows that the number of feedback in several types are significantly correlated with the assessment grades in their corresponding areas. These can be found between most types of feedback (except advice) on content and the assessment in content (|r| = .248 to .418, p < .01), two types of feedback (advice and praise) on organisation and the assessment in organisation (|r| = .223 to .450, p < .01 or .05), three types of
feedback (praise, criticism and imperative) on language and the assessment in language ($r$ = .361 to .489, $p < .01$), and three types of feedback (praise, criticism and imperative) on referencing and the assessment in referencing ($r$ = .419 to .564, $p < .01$). The results indicate that the assessment grades are correlated more with praise, criticism and imperative but less with advice, closed question and open question. They imply that teachers tended to give comments in a strong authoritative mode rather than in a less controlling mode (e.g. suggestions or hints) if a draft essay needed more changes. Given this kind of statistical summary, it would be helpful for teachers to reflect on their use of feedback types and refine their feedback practice to benefit students’ writing. Further, it would be conducive to the investigation of the impact of various feedback types on student revision in the final essay.

Table 3: Confusion matrix.

<table>
<thead>
<tr>
<th>Machine Classification</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>197</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>1722</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Correlation between the numbers of different feedback types and the assessment results.

<table>
<thead>
<tr>
<th>Feedback Types</th>
<th>Assessment Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Content</td>
</tr>
<tr>
<td>Content (Praise)</td>
<td>.418**</td>
</tr>
<tr>
<td>Content (Criticism)</td>
<td>-.293**</td>
</tr>
<tr>
<td>Content (Imperative)</td>
<td>-.315**</td>
</tr>
<tr>
<td>Content (Advice)</td>
<td>.045</td>
</tr>
<tr>
<td>Content (Closed Question)</td>
<td>-.314**</td>
</tr>
<tr>
<td>Content (Open Question)</td>
<td>-.248**</td>
</tr>
<tr>
<td>Organisation (Praise)</td>
<td>.450**</td>
</tr>
<tr>
<td>Organisation (Criticism)</td>
<td>-.126</td>
</tr>
<tr>
<td>Organisation (Imperative)</td>
<td>.014</td>
</tr>
<tr>
<td>Organisation (Advice)</td>
<td>.223*</td>
</tr>
<tr>
<td>Organisation (Closed Question)</td>
<td>.197</td>
</tr>
<tr>
<td>Organisation (Open Question)</td>
<td>.049</td>
</tr>
<tr>
<td>Language (Praise)</td>
<td>0.489**</td>
</tr>
<tr>
<td>Language (Criticism)</td>
<td>-.395**</td>
</tr>
<tr>
<td>Language (Imperative)</td>
<td>-.361**</td>
</tr>
<tr>
<td>Language (Advice)</td>
<td>-.071</td>
</tr>
<tr>
<td>Language (Closed Question)</td>
<td>-.001</td>
</tr>
<tr>
<td>Language (Open Question)</td>
<td>-.033</td>
</tr>
<tr>
<td>Referencing (Praise)</td>
<td>.419**</td>
</tr>
<tr>
<td>Referencing (Criticism)</td>
<td>-.564**</td>
</tr>
<tr>
<td>Referencing (Imperative)</td>
<td>-.535**</td>
</tr>
<tr>
<td>Referencing (Advice)</td>
<td>-.123</td>
</tr>
<tr>
<td>Referencing (Closed Question)</td>
<td>-.064</td>
</tr>
<tr>
<td>Referencing (Open Question)</td>
<td>-.194</td>
</tr>
</tbody>
</table>

$p < .05$ (2-tailed), $^{*}p < .01$ (2-tailed)
6. Conclusion and Future Work

This study proposes a method for automatic classification of teacher feedback on students’ draft essays in the EFL context. Drawing on a ten-category classification framework, the proposed method identifies types of teacher feedback by matching against syntactic and semantic features extracted from a set of manually annotated data. The findings of this study show that the proposed method could achieve very good performance in terms of classification accuracy (over 96%). They also demonstrate the potential of using the classification results as a source of reflection to enhance teachers’ feedback practice.

Future work involves designing and implementing methods for automatic identification and classification of student revision in the final essay. In addition, the association between different types of teacher feedback on student revision will be examined. This would give insights into what kind of feedback is most effective in facilitating students to make substantial and successful revisions to EFL writing.

Acknowledgements

This research was financially supported in part by the Hong Kong SAR Government under General Research Fund (GRF no. 18608816).

References

Automatic Question Generation System for English Exercise for Secondary Students

Tasanawan SOONKLANG*, Sunee PONGPINIGPINYO, Weenawadee MUANGON & Sirak KAEWJAMNONG
Department of Computing, Faculty of Science, Silpakorn University, Thailand
*soonklang_t@su.ac.th

Abstract: Automatic Question Generation (AQG) is a research trend that assists teachers to create efficiency assessments. In this paper, we propose a web-based system as a tool to generate English exercises for secondary school students automatically. The system applies NLTK library tags function words. The source texts are selected based on their grammar. The AQG module can generate exercises in ten topics; 1) Noun, 2) Pronoun, 3) Verb, 4) Adverb, 5) Adjective, 6) Comparison, 7) Conjunction, 8) Article, 9) Preposition, and 10) twelve Verb Tenses. There are four types of generated exercises; 1) complete the missing blank, 2) choose the correct answer from two choices, 3) true or false questions, and 4) error correction. The evaluation results show that our proposed system performs effectively with 97.36% F-measure. From the user experience point of view, they are well satisfied with ease of use of the system, including the capability and accuracy to generate a variety of exercises.

Keywords: Automatic Question Generation System, English Exercise, Secondary Students

1. Introduction

English language is essential for communication in non-native English countries. Therefore, it is necessary for people, especially high school students, to comprehend English by practicing. Thus, appropriate exercises are important. In their English class, students can build upon their background knowledge, clear up their confusion and improve reading comprehension by practicing a number of English exercises. The effective questioning strategies can be done using many methods. These exercises are normally created by teachers. With the Automatic Question Generation (AQG) system, exercises can be generated automatically via the computer which can produce a number of questions faster than human experts without losing the assessment quality (Pino et al 2008).

The AQG system generates reasonable questions from an input, which can be structured (for example, a database) or unstructured (for example, a text) (Susanti et al 2015). In this paper, the proposed AQG system is a web-based tool that can generate four types of questions from secondary school English textbooks. The AOG is designed for both teachers and students. The questions are automatically generated and cover ten English grammar lessons for seventh to ninth grade students in Thailand. The ten topics are 1) Noun, 2) Pronoun, 3) Verb, 4) Adverb, 5) Adjective, 6) Comparison, 7) Conjunction, 8) Article, 9) Preposition, and 10) twelve Verb Tenses. The vocabulary test, however, is not included in any topic. The question types are 1) Complete the missing blank whereby an English sentence has a gap to be filled. 2) Choose a correct answer from two choices 3) True or false questions with correction. Rather than just state their answer to be, for example, the false question, students have to correct the false question as well. 4) Error correction whereby the student must amend the incorrect word in the sentence.

The proposed AQG system is capable of checking answers, providing solutions and calculating scores to evaluate the English efficiency of students. As the text is an input, natural language processing performs an important role in our proposed AQG system.

This paper is organized as follows. In Section 2, the current scientific research in QA and QG is introduced. Section 3 describes the proposed QA system. In Section 4, the proposed framework is evaluated and explained with illustrative examples. Finally, the conclusion and future work are encompassed in Section 5.
2. Related Work

In this section, we review approaches to develop question generation on English grammar and vocabulary domain. Several researchers deployed the AQG technology as an educational application. In English education, teachers create exercises for students to learn grammatical knowledge. Chen et al (2008) presented an internet-based system that helps teachers to make cloze tests from online news articles. This system is the assisting tool for multiple-choice and fill-in-the-blank questions which allows teachers to choose the distractors from a system’s suggestion or to create them themselves.

Kunichika, et al (2001) proposed RevUP for gap-fill AQG. This system selects important sentences from texts by using sentence-ranking method from a collection of human annotations to select a gap-phrase from each sentence. Amazon Mechanical Turk is used as data for classification to predict the relevant gaps. Moreover, they use the semantic technique to choose distractors similar to the gap-phrase. Becker et al (2012) applied dictionaries of synonyms and antonyms to generate questions for grammar and reading comprehension assessment. This generates five types of questions: 1) asking about the content of one sentence, 2) antonyms and synonyms, 3) modifiers appearing in plural sentences, 4) asking about the contents represented by plural sentences with relative pronouns, and 5) asking about time and space relationship. This research used syntactic and semantic techniques to extract information from an original text. They use syntactic trees to label parts of speech and modify word relationship. Moreover, Becker et al’s system employs the feature structures to extract grammar. Flesch (2016) proposed generating questions from text, which helps children in grade 1-3 to understand contents. The system generates questions from the situation model of text, which is constructed by schema-building rules. The situation model is an intelligent method to be used with the appropriate mental age required to transform the sentence into a question.

Brown et al (2005) proposed the REAP system that generates many types of vocabulary questions, including definition, synonym, antonym, hypernym, hyponym, and cloze questions, by using WordNet (Fellbaum, 1998). Gap-fill questions are generated with three stages: sentence selection, key selection and distractor selection. Hoshino and Nakagawa (2007) presented gap-fill questions by choosing the informative sentences from the documents. Moreover, they applied syntactic and lexical features in the process of distractor selection. Narendra et al (2013) presented automatically generating English vocabulary tests. The method consists of four components: target word, reading passage, correct answer and distractor. The target words are taken from web texts and used in reading passages. Then both correct answers and distractors are generated from WordNet lexical dictionary. Kumar et al (2015) presented an automatic cloze question generation system that generates a list of important cloze questions from English articles. They proposed a semi-structured approach. Firstly, for example, knowledge from a Cricket portal is extracted as the summarized results. Then, the top-ten-ranking sentences are selected. Lastly, the meaningful distractors are chosen from the knowledge base.

3. Methodology

3.1 System Architecture

This web-based application is designed for both teachers and students. Teachers can use the system to create the English grammar exercises for students in secondary school. The questions can be generated automatically from source English texts. The topics covered ten English grammar lessons for seventh grade to ninth grade in Thailand. Our system supports four types of question. Table 1 describes lists of English topics and our generated question types.
Table 1: Lists of English topics and our generated question types.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Fill in the blank</th>
<th>Either/Or choices</th>
<th>True/False</th>
<th>Error Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preposition</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verb Tense</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noun</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pronoun</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Verb</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Adverb</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjective</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Conjunction</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since teachers can select text from any source as an input, such as news, documents, reading passages and so on, our system will firstly analyze the suitability of the input text for the students in terms of their reading level. We use the Flesch Reading Ease test (Flesch, 2016) for calculating the readability scores. The formula is as follows:

\[
\text{ReadingEase} = 206.835 - 1.015 \left( \frac{\text{TotalWords}}{\text{TotalSentences}} \right) - 84.6 \left( \frac{\text{TotalSyllables}}{\text{TotalWords}} \right) \tag{1}
\]

We also used CMU Pronouncing Dictionary (CMUdict, 2016) to provide words and their phonemes to compute the total syllables by counting the vowel phonemes of all words. This test rates text on a 100-point scale. The high score means the document is easier to understand. The appropriate score for our students’ level is between 60 and 69.

After choosing the proper text, the teacher will select the topics and the question types to build the exercises. Then, the system generates all possible questions corresponding to the selected topics. It is possible that one sentence can create more than one question for the same and different topics. For example, for the input sentence “He sat down and ate breakfast”, the system will generate two questions of verb tense topic. The first one is “He ___ down and ate breakfast. [sit]” and the second one is “He sat down and ___ breakfast. [eat]”. Therefore, the questions will be approved and marked by the teacher to form the final exercises.

3.2 System Architecture

There are several steps to generate questions and answers. Firstly, source input text is tokenized into sentences and words. Secondly, part of speech (POS) tag is labeled to each word. After that, the sentence with the tags for the target topic will be searched. Then, a question can be created with its correct answer. Finally, incorrect choices can also be generated, if necessary. Figure 1 describes the processing steps to generate questions and answers.

In our implementation, we used NLTK library for tokenizing and tagging. For example, the sentence “The best time to fish is early or late in the day” is transformed with tags as “The<DT>, best<NN>, time<NN>, to<TO>, fish<NN>, is<MD>, early<JJ>, or<CC>, late<JJ>, in<IN>, the<DT>, day<NN>”. The Penn Treebank tag set is used for labeling tags.
In the searching process, we put "#" to cover the word with a target tag in the sentence. For example, if we need a sentence with a coordinating conjunction, the example sentence will be “The best time to fish is early #or# late in the day.” Then, we store “or” as the answer. After that, we create an incorrect choice. There are different techniques based on each topic. In our example for a coordinating conjunction, we randomly select a distractor from the list of vocabularies with a conjunction tag <CC> and a different meaning from the correct word. In this case, the word “and” is given as a distractor.

Here is the example question of all four question types from the input sentences.

**Topic: Verb tense (past tense)**
Input sentence: He sat down and ate breakfast.
Fill in the blank: He ___ down and ate breakfast.

**Topic: adjective**
Input sentence: It was late at night.
Either/Or choice: It was ____ (late/lately) at night.
True or false: It was lately at night.
Error correction: It was lately ____ at night.

**Topic: Pronoun**
Input sentence: They are both considered stable production releases.
Error correction: Their ____ are both considered stable production releases.

### 4. Evaluation

#### 4.1 Evaluation Results

To evaluate the English exercises that are created from our system, we used 30 reading passages from English textbooks taught in Thailand secondary schools as input text. We measured the performance of the system by using precision, recall and f-measure. Precision is a measure of how precise is the system in generating candidate questions that correctly match with the selected topics. Recall is a measure of how many truly relevant sentences of the chosen topics are returned. F-Measure is the harmonic mean of precision and recall. Let \( A \) is the number of sentences that the system selects and matches correctly with the topics. Let \( B \) is the number of sentences that the system selects but which are mismatched with the topics. Let \( C \) is the number of sentences in a lesson which match the chosen topics, but they are not selected by the system. The precision, recall and f-measure are defined as follows.

\[
Precision = \frac{A}{A + B} \tag{2}
\]

\[
Recall = \frac{A}{A + C} \tag{3}
\]

\[
F - \text{Measure} = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{4}
\]
The percentage of precision, recall and f-measure of our experiments of each English topic and the average percentage of the system performance are shown in the table 2.

Table 2: Percentage of precision, recall and f-measure of all topics.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Percentage</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>Preposition</td>
<td>99.67</td>
<td>100.00</td>
<td>99.83</td>
<td></td>
</tr>
<tr>
<td>Verb tenses</td>
<td>97.95</td>
<td>75.85</td>
<td>84.63</td>
<td></td>
</tr>
<tr>
<td>Noun</td>
<td>98.78</td>
<td>99.13</td>
<td>98.95</td>
<td></td>
</tr>
<tr>
<td>Pronoun</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>Verb</td>
<td>97.11</td>
<td>95.00</td>
<td>96.04</td>
<td></td>
</tr>
<tr>
<td>Adverb</td>
<td>99.05</td>
<td>100.00</td>
<td>99.52</td>
<td></td>
</tr>
<tr>
<td>Adjective</td>
<td>99.05</td>
<td>100.00</td>
<td>99.52</td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>98.89</td>
<td>100.00</td>
<td>99.49</td>
<td></td>
</tr>
<tr>
<td>Conjunction</td>
<td>100.00</td>
<td>91.67</td>
<td>95.65</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>99.05</td>
<td>96.17</td>
<td>97.36</td>
<td></td>
</tr>
</tbody>
</table>

The user experience of the system is evaluated by ten users who generated exercises and then tested these exercises with ten secondary school students. The result shows that they are well satisfied with the ease of use of the system, the variety of question types, and the capability to select the proper sentences to generate questions.

4.2 Result Analysis

The above results demonstrate satisfactory performance. However, the verb topic has the lowest value in precision score. Also, the twelve verb tenses category has the lowest percentage of recall. After carefully analyzing the results, we found that the lowest precision arises from faulty NLTK tagging. For example, in our case the words that suffix with –ing have to be a noun but NLTK tagged these as a verb. Moreover, some selected sentences do not match with the correct syntax, such as the sentence: “They are fun to play.” The NLTK library tags “fun” as <VBN> refers to the verb (past participle) but it actually serves as an adjective. In another example, “We never hear police sirens or fire sirens”, the extracting system tagged the grammar of the sentence as follows.

We<PRP>, never<BB>, hear<JJ>, police<NNS>, sirens<NNS>, or<CC>, fire<VB>, sirens<NNS>

The false result of NLTK tagging function is the word “hear” <JJ> as the adjective, but in fact "hear" is a verb. Thus this sentence is not selected to generate question. This is the reason that the scores of precision, recall and f-measure were decreased.

Because the NLTK tagging function provided the wrong tag for sentences and made the sentences inappropriate to the verb topic. In order to solve this problem, we analyzed all the mistakes of NLTK including the structure of the sentence, position of the word and so on. It was noticed that the tag type were often wrong in respect of the received results. For example, NLTK tags possessive pronouns as the plural noun <NNS>. Thus, when the user selects the topic for pronoun grammar to generate questions, these sentences are not used. In this case, we modified this by changing the tag to personal pronoun <PRP>.
5. Conclusions

The proposed system AQG was used to generate English Exercise for secondary school students. It is a Web-based application. The developed system applies NLTK library tags function words, and source sentences are selected based on grammar. The experiment we conducted generated exercises from 30 reading passages which showed that our system performed well with 97.36% F-Measure. Word tagging enhances the system to select correct word function and structure of sentences. However, one drawback of using the NLTK library is that this library tool does not tag correct words completely. Therefore, we plan to apply rules and templates to enhance our system performance. In addition, we also plan to develop the system to cover lessons for higher level students.

References


Student Engagement with an Online Pre-enrolment English Course at a Japanese University

Adam SMITH & Andrew JOHNSON
Future University, Japan
*adam@fun.ac.jp

Abstract: The authors manage an English language course for high school students who successfully apply for early enrolment to a Japanese university each year. The course is non-credit bearing and although encouraged, participation is at the students’ volition. Its primary aim is to help students maintain their English language skills in the four or five months before university lectures start in early April. In keeping with the university's practice of utilising technology in the classroom, the majority of the course is online, using the open-source Moodle LMS. The online portion of the course includes interaction with their future classmates and upperclassmen (currently enrolled at the university) in asynchronous forum discussions. In this way, too, students create their own course content. The success of such a course depends on the level of student activity, and as such the authors have an interest in maximising this. This paper will describe the key features of the 2017 iteration of the course and the extent to which students engaged with it. The results of a post-course student survey will be analysed to understand students' opinions of the course, and the main impediments to student engagement. Based on these, some possible changes for the 2018 version will be described.

Keywords: Moodle, Japan, English, course design

1. Background

1.1 The University Enrolment Process in Japan

In Japan, approximately three quarters of high school students enter a university or other higher education institution after graduation (MEXT, 2011, p.10). There are three main streams through which high school students enrol in universities. The most commonly used one is the General Entrance stream, held in January and February, with results published less than two months before the academic year starts in April. Most universities also offer places to a limited number of students through one or two alternative streams. These are the “Suisen” (Recommendation) and the AO (Admissions Office) entrance streams. In these, students are given an opportunity to argue their suitability for an institution on the basis of their academic performance, aptitude and other achievements.

Universities typically publish the results of the AO and Suisen exams in the period from mid-November to mid-December, however some universities release them as early as September (Iwate Prefectural University, 2017). While early acceptance creates significant security for both the university and student, this situation can have wide-ranging negative impacts. Many high schools have a strong focus on university entrance and have classes devoted to exam preparation. With no entrance exams to prepare for, students who have been accepted early lose their primary motivation to study, and they can distract or disturb their classmates who are still preparing for exams. Unless their teachers create additional activities, accepted students have little motivation to study, which can result in decreased study skills and study habits. Students who have an extended break from study can experience a difficult transition to university life, with further flow-on effects for their new classmates.

In order to maintain student motivation and allow for a smooth transition to tertiary studies, some Japanese universities provide students who pass their Suisen and AO entrance exams with a supplementary program of study during the last months of high school, called a pre-enrolment course.
1.2 The Setting

The authors’ university offers 4-year courses in fields relating to computer science and information design. There is a strong emphasis on the use of technology for learning. Four courses are fully online and an increasing number of classes have a blended learning format. English is not a major, however there are compulsory English language courses, and English is used as a medium of instruction and assessment in first- and second-year Communication classes. Thus, students are expected to have at least a basic level of English language skills.

The university holds Suisen and AO entrance exams each year, and the results are published by early December. Approximately 100 students are offered a place, which is equivalent to 40% of the annual first-year student intake. Pre-enrolment courses in English and mathematics have been offered to students since 2003, and the authors have been responsible for the design and management of the English course since 2014.

1.3 Course Format

When designing the pre-enrolment English course in 2014, the authors considered three formats: on-campus, paper-based and online. In an on-campus course, students gather at the institution for a series of events with teachers. A paper-based course involves sending students written material which, upon completion, is sent to the teacher for grading and then returned to the students with feedback. An online course uses a learning management system (LMS) to provide students with materials and activities. Smith et al. (2017) discuss the benefits and drawbacks of each in relation to the authors’ university.

The most attractive option was the online format, as all of the benefits were in alignment with the university’s philosophy. The drawbacks, though not insignificant, were considered manageable. The decade to 2016 saw significant increases in both internet penetration (Internet Live Stats, 2017), and smart device ownership in Japan. (eMarketer, 2017). While online courses require some ICT skills, this was seen as a good opportunity for students to develop these skills, which will be used regularly from April. For these reasons, the authors chose the online format.

1.4 The Students

First-year students at the authors’ university have the following general characteristics: low motivation to study English, low English language skills, and little-to-no experience with e-learning. Each July, first-year students take the TOEIC Bridge test, which is designed for false beginners of English. Data for the 2017 cohort was unavailable at the time of writing, however the 2016 cohort’s mean grade was 134 out of 180, which is equivalent to a TOEIC score of approximately 370 (ETS, 2006) or a CEFR level of A2 (ETS, 2015). Averages for previous years were also similar.

In early December 2016, a total of 98 students had been offered places at the authors’ university resulting from the AO and Suisen entrance exams. They were made aware of the pre-enrolment course and all said that they would participate in it.

2. The 2017 Pre-enrolment English Course

2.1 Aims

Although the primary focus of the course is English, it is designed to provide students with a multifaceted learning experience. Thus, the aims of the course are to provide opportunities to:

- improve students’ English language skills, in particular reading and writing;
- experience e-learning and develop their ICT skills;
- communicate with their future classmates; and
- learn about and prepare for university life.
2.2 Course Design

The course is non-credit bearing and participation is voluntary, so care was taken to make it accessible and attractive to students. Instructions and explanations mailed to students were bilingual, and students could choose their preferred language in the Moodle course.

It was suggested to students that they access the course for at least two hours a week. The authors were mindful of the likelihood that many students would access the course less frequently, and made all content and activities "stand alone" and available until the end of the course.

In recent years there has been an increase in students accessing the course using a smartphone or tablet as the main device. In addition to choosing a Moodle theme designed to adapt to smaller devices, a number of modifications to the course design were made in order to accommodate mobile use, such as making images and PDF files responsive, which makes them easier to read on small screens.

The course included a number of resources and activities, including the following:

- A self-introduction forum. As the first forum in the course, it contained more detailed instructions on how to create a post and how to add media such as photos.
- Weekly discussion forums. These were designed for students to communicate with each other in English and to learn more about their classmates. The topics were designed to elicit student ideas and opinions, with some having an emphasis on academic reflection.
- Reading activities. In January a book was sent to students: Oxford Bookworms' *The Coldest Place on Earth*. This is an illustrated Stage 1 graded reader about early Antarctic exploration. This was supplemented with online quizzes.
- A virtual exchange. A short virtual exchange with current second-year students took place, in which they described their everyday university lifestyles using text and photographs.
- English grammar resources. Access to an in-house online language learning resource called English Foundations was provided for the students.
- Information about the university. This was a series of photos and explanations written in English by a student who assisted with administering the course.
- An open forum. Students could discuss any topic they wished to in Japanese.

As can be seen, the pre-enrolment course offered a combination of communicative activities and self-study opportunities.

3. How Students Engaged with the Course

As a way of assessing the course's effectiveness, the authors examined the level of student engagement to identify issues related to student engagement. By pairing these findings with survey results (Section 4), a more complete understanding of the student experience is possible.

3.1 The Moodle course

As mentioned previously, one of the benefits of an online course over a paper-based course is the ability to get an understanding of student activity by reviewing the course logs. Thus, the authors could determine how often the students accessed the course and what kinds of activities they engaged in.

The 2017 cohort consisted of 98 students. It was recommended that they access and participate in the course for at least two hours a week during the 14 weeks it was running. If a student did this, they would have accessed the course on a minimum of 14 separate days. Figure 1 below shows the number of days each student accessed the Moodle course. Nineteen students, referred to as high-access students, accessed the course on 14 or more individual days, which means that weekly access was possible. A total of 77 students accessed it on less than 14 days; they are referred to as low-access students. Two did not access the Moodle course at all.
The discussion forums formed the main part of the course. They were added to provide an environment that is useful for both active students as well as those who prefer just to read (Wegener et al., 2002). Table 1 below shows the average number of posts written to the forums and the average number of forum-post views by students in each access group. Data from the self introduction forum, which was the most popular, is also included. It is clear that the level of participation in the discussion forums is quite low. In discussion forums students create the content, and if not enough is written then there is not enough to satisfy the readers.

<table>
<thead>
<tr>
<th>Student group</th>
<th>Forum posts</th>
<th>Forum views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-access group</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>High-access group</td>
<td>3</td>
<td>89</td>
</tr>
</tbody>
</table>

4. Factors Affecting Students Accessing the Course

The course closed on the day of the entrance ceremony in early April. During the following Orientation Week all 98 students completed an online survey about the course. One of the questions in the end-of-course survey was "Why did you not access the course more often?" (in Japanese). They could select up to three main reasons from a list which included "I accessed the course enough". The most popular primary reasons are shown in Table 2 below.

<table>
<thead>
<tr>
<th>Student group</th>
<th>#1 1st reason</th>
<th>#2 1st reason</th>
<th>#3 1st reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-access (n=21)</td>
<td>I did not know what to do in the Moodle course (6)</td>
<td>I did not have regular access to a computer or device (4)</td>
<td>The contents of the course were not interesting (3)</td>
</tr>
<tr>
<td>Low-access (n=75)</td>
<td>I did not know what to do in the Moodle course (24)</td>
<td>I did not have regular access to a computer or device (23)</td>
<td>I studied English somewhere else (9)</td>
</tr>
</tbody>
</table>

It should be noted that although two hours a week was suggested, it may not have been necessary for all students to access the course this much. It is understood that some students were engaged in activities which helped them to maintain their English language skills during this time, however in order for students to take advantage of the other potential benefits it was suggested that they get involved in the course anyway.

Nearly a third of students who accessed the course reported that not knowing what to do was the primary reason for not accessing it more. There are several possible explanations for this. First, while
the authors deemed that the instructions and explanations (in both Japanese and English) were sufficient, they might not have been clear for some students. Secondly, Japanese high school students typically are not used to activities such as forum discussions that do not have a clear ending point. Third, some students might not have had the ICT skills to successfully follow the instructions. Finally, some students might have felt reluctant to engage with an online environment which was out of their comfort zone.

5. Student Opinions of the Course

One of the questions asked students to indicate the three things they liked most about the course. The ten most popular choices are shown below in Figure 2.

![Figure 2. What students liked about the course](image)

In spite of the abovementioned issues, the students as a whole liked e-learning, as well as using English. In addition, the high-access students especially liked the English language aspects of the course, such as reading the forums and the book. In contrast, the low-access students favoured the course's communicative activities, such as learning about the university and having contact with future classmates.

6. Conclusion

The authors have created an English course for high school students who have been accepted to university up to five months before the start of the academic year. It is essentially a fully online course, which, although it has some benefits for both students and course administrators, it has a few challenges. Since the course is non-credit-bearing, student participation is solely based on internal motivation.

Students who accessed the course did participate in the core activities, however to a lesser extent than was suggested to them. In order to improve this, the authors are considering reducing the frequency of forums from weekly to fortnightly, and introducing some scaffolding activities at the start of the course to guide the students into an online learning environment. In addition, attention will be given to the format of the course, so that it is easily accessible by mobile device.
References

Captioning Methods of Lecture Videos for Learning in English

Veri FERDIANSYAH & Seiichi NAKAGAWA*
Toyohashi University of Technology, Japan
*nakagawa@tut.jp

Abstract: This paper describes a study on the effect of using captions in lecture videos taught in English to assist non-native English speakers in learning. Lecture videos used in this paper were taken from MIT OpenCourseWare. We divided captions into 5 different types: (1) full text captions, (2) important sentence captions, (3) important phrase captions, (4) important keyword captions, and (5) no captions. Each caption was created manually and embedded in the videos. We prepared captions in two languages; English captions for foreign students and Japanese students, and Japanese captions for Japanese students. Participants for the experiment were chosen randomly from non-English native speaking countries and each participant must follow a set of tests. Experiment results showed that captions (2) or (3) helped participant the most in understanding the content of lectures.

Keywords: lecture speech, full text captions, important sentence captions, important phrase captions, important keyword captions, no captions, English captions, Japanese captions

1. Introduction

The widespread of Internet and the openness of education has made learning more accessible to everybody. Massachusetts Institute of Technology (MIT), with its MIT OpenCourseWare site, is the pioneer of the open courseware movement. Many other universities have followed the open courseware movement and opened their educational materials (e.g., course textbooks, presentations, videos, assessments, simulations, etc.) that are either (a) licensed under an open copyright license (e.g. Creative Commons) or (b) in the public domain so everybody can gain access to the resources at no costs. However, many of those educational materials, particularly lecture videos, are available without any captions or separate transcriptions. Therefore, for example, lectures in English, for students in non-English speaking countries such as Japan, will be very hard to understand the content of the lectures without getting any help from either captions or transcriptions.

The purpose of this study is to see whether captioning lecture videos in English will benefit students for learning in foreign language while investigating which type of caption will give more benefit to the learners. We divided captions into five different types and each caption was created manually by the authors and embedded in the videos using video editing software. We randomly chose participants for the experiment which are students from non-English speaking countries and each participant must follow a set of tests. During the tests, each participant will be shown 5 different videos with 5 different captions. This study showed that full text captions and important sentence/phrases captions helped participant the most in understanding the content of lectures. We also prepared the same five different types of captions in Japanese for Japanese students.

This paper is organized into the following five sections. In Section 2, previous works related to this study is explained. Methodologies are described in Section 3. Section 4 presented the results of our experiments and in Section 5, a conclusion is made and we explained our future work (Ferdiansyah and Nakagawa, 2013).

2. Related Work

Guillory et al. (1998) tested a keyword captioning method based on a hypothesis that keyword captions, while presenting learners with lesser material to read, didn’t weaken their comprehension in
the spoken message. Zanon et al. (2006) distinguished three types of subtitling; (1) bimodal subtitling from English dialogues to English subtitles, (2) standard subtitling from English dialogues to subtitles in the learner’s mother tongue, and (3) reversed subtitling from dialogues in the learner’s mother tongue to English subtitles. Those subtitling types were then used by Zarei (2009) where the author investigated the effect of bimodal, standard, and reversed subtitling on L2 vocabulary recognition and recall. Winke et al. (2010) attempted to investigate the effects of captioning during video-based listening activities for second language learners.

Recently, Perez et al. (2014) also compared three types of captions: full captions, keyword only captions and no captions. They showed that the full caption outperformed the keyword only caption and no caption on the global comprehension questions, on the other hard, there was no difference on the detailed comprehension questions. Yang and Chang (2014) compared full captions, keyword only captions and annotated keyword captions for reduced forms such as the processes of contraction, elision, assimilation and reduction. Lwo and Lin (2016) compared four different types of captions; no captions, Chinese captions, English captions and Chinese plus English captions. They showed that the effect of different captions varied depending on student’s proficiency levels.

Our purpose is to investigate how to present captions or translated captions by English and Japanese. Our final goal is to present the speech translated captions automatically based on this study. The following comparative experiments had been done in 2013, and then we extend the consideration and discussion.

3. Methodology

We used two methodologies for this study: (1) using English captions for targeting both foreign students and Japanese students and (2) using Japanese captions for targeting only Japanese students. The preparations and experiments conducted for both methodologies are the same and explained in the following sub-sections.

3.1 Video Preparations

In our experiment, we used videos from MIT OpenCourseWare lectures on computer science fields (Abelson, 2008). While screening and selecting the video materials that would be used, several criteria were maintained. The first and the most important criterion was that the video materials must have a transcription for our experiments, because manually transcribed videos needed are very impractical and time consuming. Thus, the pre-defined transcription is very important. The topics of the videos are also important criteria. Many topics are available but not all of them are common topics in other universities. It has to be a general topic, taught in every Computer Science department in any universities. With these criteria in mind, the lecture for video materials was selected.

From over 200 courses available in Electrical Engineering and Computer Science Department in MIT OpenCourseWare, one course was selected: Introduction to Computer Science and Programming (Fall 2008). From this course, we selected five different lectures to be used as our video materials. Once the video test materials were chosen, a shorter video about five minutes in duration was made from the original 60 minutes video for each lecture.

3.2 Captions

The chosen lecture video comes with a separate transcription. We utilized this transcription to create our captions. We created five different caption types manually: (1) full text captions, (2) important sentence captions, (3) important phrase captions, (4) important keyword captions, and (5) no captions.

To differ it from the full text captions, we limited the number of sentences in important sentence captions to be about 40-50% from the total number of sentences in the full text captions. Important phrase captions were created from the important sentence captions. From each sentence in the important sentence captions, we extracted 1 or 2 important phrases while maintaining the number of words in the important phrase caption is about 40-50% from number of words in the important sentence captions. As for important keyword captions, it was extracted from the important phrase captions. From each important phrase, we extracted about 1 or 2 important keywords. Using the
captions, we embedded the captions into the lecture videos. The final outcome was five videos with five different types of captions in English for each lecture.

For the Japanese captions, we also created the same five types of captions as the English version. Each caption type was a direct translation by a professional translator from its corresponding English caption.

### 3.3 Test set

We created a test set consisted of three different types of test to find out whether captioning video really helps learners to understand the content of the video and to measure which type of captions gives the most benefit to the learners: (1) pre-comprehension objective test, (2) comprehension objective test, and (3) caption helpfulness and understandability subjective test.

All questions and possible answers in all tests were originally written in English. For English captions experiment, in order to accommodate Japanese students who may have difficulties in reading English questions, we provided Japanese translation below the English questions but all possible answers were left un-translated in purpose. For Japanese captions experiment, all questions and possible answers are translated into Japanese language.

### 3.4 Scoring

Scoring method for the pre-comprehension and the comprehension test was different from the scoring method for caption helpfulness and understandability test. For the pre-comprehension and comprehension test, 10 points were given for each correct answer and 0 point for wrong answer. For caption helpfulness and understandability test, the score depends on participants’ answer about each caption’s level of helpfulness and understandability from one to five.

### 3.5 Participant Selection

Participant was divided into two different groups. The first group is international foreign students (FS) and the second group is Japanese students (JS). Several criteria must be met by each participant: (1) they have to be majored in Computer Science in Toyohashi University of Technology, (2) for international students, they have to come from non-English native speaking country, and (3) for Japanese students, they must have a TOEIC score not less than 400 (420 for TOEFL). A total of 21 subjects – 10 international students and 11 Japanese students participated in the testing sessions. Additionally, we recruited other 10 Japanese students as the third group who evaluated only five caption types of Japanese. The TOEIC score for some of them is less than 400 (420 for TOEFL). In order to make it comparable, we converted Japanese students’ TOEIC score into TOEFL score by using the converting equation given by “TOEFL score = 0.348 × TOEIC score + 296”.

For references, we investigated the relationship between TOEFL score and correct word transcription rate (WTR) (Goto et al., 2016). The new 15 Japanese students were evaluated by transcribing while listening to English lectures. We allowed to listen them three times, but the transcribing time was limited to less than 150 seconds per utterance. The correct word transcription rate (WTR) was about 32% for students with average TOEFL score of 461 (TOEIC score of 475), 38% for 504 (TOEIC score of 600), and 42% for 556 (TOEIC score of 750). This shows that it is very difficult to listen or understand MIT English lectures for Japanese students. On the other hand, the automatic speech recognition rate by using our speech recognizer (Goto et al., 2016) which is better than students’ transcription (the word correct rate is about 80%).

### 3.6 Procedure

All testing sessions were conducted in our Spoken Language Processing laboratory. All sessions were administered by the first author. Each testing session lasted approximately one hour.

The participants firstly took pre-comprehension test. There was no time limit for this test. They watched a series of five videos with the five different types of captions in randomized order. While watching the videos, the participants were not allowed to take notes, pause, rewind, and fast
forward the video. Each participant viewed all five videos consecutively without any break. After watching all videos, all participants were required to take comprehension test and caption’s helpfulness and understandability test. There was also no time limit for both tests.

4. Results

Figure 1 presents the comparison among the pre-comprehension test score, comprehension test score, and TOEFL score for each participant who viewed the English captioned videos. The data showed that 19 participants had a higher score in their comprehension test if compared to the pre-comprehension test. Only 2 Japanese students had lower score in their comprehension test compared to the pre-comprehension test.

Figure 2 presents the comparison among the pre-comprehension test score, comprehension test score, and TOEIC score for each participant who viewed the Japanese captioned videos. The data showed that all participants had a higher score in their comprehension test if compared to the pre-comprehension test.

Figure 3 shows the average subjective evaluation score for caption helpfulness and understandability test. The result shows that the full text caption was the most helpful and the most understandable type of caption according to the participants, followed by the important sentence captions, important phrase captions, important keyword captions and the least helpful and the least understandable was no caption. There were no differences among the full text captions, the important sentence captions and the important phrase captions in English captions. However there was large difference between the full text captions or the important sentences and the important phrase captions in Japanese captions translated from English. We can conclude that the useful caption type depends on the learner’s language ability. On the other hand, not all participants learned towards full text captions on their comments in the open question about what did thought about each type of caption.

![Figure 1. Pre-comprehension test vs. comprehension test vs. TOEFL score for English captioned videos](image_url)
Figure 2. Pre/post-test and TOEFL scores for Japanese captioned videos

Figure 3. Test score for subjective evaluation

Figure 4. Captions Score for objective evaluation
Some participants’ comments are: “Other captions than full text captions can be misleading.” (foreign student having 550 TOEFL score); “Full text captions are the most helpful.” (foreign student having 610 TOEFL score); “I prefer full text captions especially if the topic is hard so I can focus on the captions.” (foreign student having 600 TOEFL score); “The speed of full text captions is too fast! I think the important sentence captions are more suitable for me.” (Japanese student having 503 TOEFL score (595 TOEIC score)); “Important sentence captions are easier to understand than full text captions.” (Japanese student having 558 TOEFL score (690 TOEIC score); “Important phrase is good but still lacking information.” (Japanese student having 487 TOEFL score (550 TOEIC score).

Figure 4 shows the average objective evaluation score for each type of caption. This score was calculated by grouping comprehension test questions based on caption type. Each caption has 2 questions in the comprehension test and we calculated the average of each caption type. The important phrase caption was the most useful. From Figures 3 and 4, we can expect that speech translation for only important sentences or important phrases is enough as understandable captions for Japanese students, especially, who have not good English ability.

5. Conclusion and Future Work

In this paper, we described a study of the effect of captioning lecture videos for learning in foreign language. We compared five different types of captions in English and Japanese for English lectures. We showed that captioning lecture videos benefited the students for learning in foreign language. From the caption helpfulness and understandability test results, we showed that subjectively, the full text captions and important sentence captions are the most helpful and understandable caption types but objectively by looking at the captions score, the important phrase captions was the most useful caption types.

We have been developing automatic Japanese captioning system for English lectures (Goto et al., 2016). Firstly, we create the ASR system for English lectures that we use to transcribe all lectures from MIT OpenCourseWare. Then we will create the captions, especially important sentence and important phrase captions by using the transcribed lecture while translating it into Japanese language.

Acknowledgment

This work was supported by JPS KAKENHI Great Number 2528002 (2013-2017).

References

Perez, M. M. et al. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension, ReCall, 26(1), 21-43

907
Using Learning Analytics to Support Computer-Assisted Language Learning

Huiyong LI*, Hiroaki OGATAb, Tomoyuki TSUCHIYAc, Yubun SUZUKIc, Satoru UCHIDAc, Hiroshi OUSHId & Shin’ichi KONOMId

*Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan
cFaculty of Languages and Cultures, Kyushu University, Japan
dFaculty of Arts and Science, Kyushu University, Japan

*lihuiyong123@gmail.com

Abstract: Computer-assisted language learning (CALL) is often used as an approach to foreign language teaching and learning in higher education. The CALL course is widely offered at universities in Japan to allow freshman students to perform self-regulated learning (SRL) with e-learning materials to develop language skills in grammar, listening, and reading. SRL is an active learning process used to regulate and monitor learning cognition, motivation, and behavior, thereby setting personal goals (Wolters, Pintrich & Karabenick, 2005). In the process of SRL, freshman students often do not recognize how they are learning and thus do not appreciate many beneficial learning strategies (Bjork, Dunlosky, & Kornell, 2013). The learning behavioral differences between novice and skillful self-regulated learners reveal that novice self-regulated learners avoid self-evaluation and have negative self-reactions (Schunk & Zimmerman, 1998). The students with low SRL skills are more likely to obtain lower achievement, and therefore it is critical to provide adaptive support to improve freshman students' SRL skills.

In addition, it is difficult for instructors to grasp students’ learning situation due to the large amount of evaluation work. Since a large number of students and online learning materials need be paid attention to, it is limited for instructors to provide individual instruction to students in the CALL environment.

Learning analytics studies could be applied to improve education, create learning supports, establish learning models, and so on (Gray, 2014). Using learning analytics and the massive learning logs in current CALL course, it is able to support instructors and students efficiently. In this research, the learning logs were collected and extracted from the server of the CALL course, and time-based indicators from students' log data were analyzed to understand the learning activities. Utilizing time-based indicators, a learning support system on the CALL course is proposed. The aims of the system are to support students’ self-monitoring and instructors’ decision making easily.

Keywords: Learning analytics, self-regulated learning, computer-assisted language learning

1. Introduction

Computer-assisted language learning (CALL) is often used as an approach to foreign language teaching and learning in higher education. The CALL course is widely offered at universities in Japan to allow freshman students to perform self-regulated learning (SRL) with e-learning materials to develop language skills in grammar, listening, and reading.

SRL is an active learning process used to regulate and monitor learning cognition, motivation, and behavior, thereby setting personal goals (Wolters, Pintrich & Karabenick, 2005). In the process of SRL, freshman students often do not recognize how they are learning and thus do not appreciate many beneficial learning strategies (Bjork, Dunlosky, & Kornell, 2013). The learning behavioral differences between novice and skillful self-regulated learners reveal that novice self-regulated learners avoid self-evaluation and have negative self-reactions (Schunk & Zimmerman, 1998). The students with low SRL skills are more likely to obtain lower achievement, and therefore it is critical to provide adaptive support to improve freshman students' SRL skills.

In addition, it is difficult for instructors to grasp students’ learning situation due to the large amount of evaluation work. Since a large number of students and online learning materials need be paid attention to, it is limited for instructors to provide individual instruction to students in the CALL environment.

Learning analytics studies could be applied to improve education, create learning supports, establish learning models, and so on (Gray, 2014). Using learning analytics and the massive learning logs in current CALL course, it is able to support instructors and students efficiently. In this research, the learning logs were collected and extracted from the server of the CALL course, and time-based indicators from students' log data were analyzed to understand the learning activities. Utilizing time-based indicators, a learning support system on the CALL course is proposed. The aims of the system are to support students’ self-monitoring and instructors’ decision making easily.
2. Related Work

2.1 SRL in a Computer-based Learning Environment

SRL is defined as an active and constructive process through which learners can set goals, and monitor and control their cognition, motivation, and behavior (Pintrich, 2000). It is also characterized as a self-directive process as self-beliefs enable learners to transform their academic abilities (Zimmerman, 2008). Winne and Hadwin (1998) proposed that SRL included four phases: defining the task, setting goals and plans, enacting tactics, and adapting metacognition. Therefore, learners need to analyze the learning context and define tasks, set the appropriate learning goals and make plans, select the effective learning strategies to use, monitor the whole learning process, and evaluate their learning performance.

Previous studies indicated that SRL is the crucial skill for success in computer-based learning environments (Adeyinka & Mutula, 2010). However, learners cannot always regulate themselves successfully because of reasons such as lack of good strategy use, lack of metacognitive knowledge, failure to control of metacognitive processes, or lack of experience in learning environments with multiple representations. Thus, how to foster SRL ability has become a central issue in the field of education research and practice.

In order to support SRL in the CALL course, instruments that capture students' self-regulation are critical. Most studies on self-regulated learning have used self-report instruments, which not only are intrusive but also are limited to capturing actual self-regulated behaviors in learning contexts. However, this issue can be resolved via online trace data use, and such technologically mediated learning environments enable the collection of a comprehensive set of student learning behaviors that occur in learning environments (Pardo, 2014).

2.2 Learning Analytics

Learning analytics are driven by the collection and analysis of traces that learners leave behind (Greller & Drachsler, 2012). It can help to understand and optimise the learning process and the environments in which it occurs (Siemens & Long, 2011; Ogata & Mouri, 2015). Until now, learning analytics are mostly feedback to the users in web-based learning dashboards (Verbert et al., 2014). Those dashboards can support raising awareness and reflection of individual and peer performance, suggest additional learning activities or content and therefore can have an impact on the learning behavior. For instance, monitoring the state in a learning activity can motivate the learner towards the accomplishment of a learning goal. This cognitive process has been defined as “self monitoring”, and “understanding how to learn” (Candy, 1991).

However, there were few research conducted using massive learning data (2499 students’ learning data) in the field of learning analytics. Further, considering the fact that students participating in online learning exhibit a lack of time management regarding self-regulated learning, such as cramming and procrastinating, and it is therefore critical to provide individual feedback to help students take proactive actions. In this research, the massive learning data were analyzed in order to understand the actual behaviors of all students, and then a learning support system was proposed through the learning dashboard in order to provide prompt feedback to support students’ self-monitoring and instructors’ decision making.

3. Research Methods

3.1 Setting and Participants

A total of 2,631 students enrolled in 50 CALL classes at Kyushu University, our institution. The CALL classes were provided for freshman students from all departments with two credits from spring and fall semesters 2016. The students were supposed to perform self-paced language learning outside of the classroom. Table 1 shows the course schedule in the spring semester 2016. To increase the motivation of the students, four sub-deadlines were set in one semester. The students were required to complete the assigned materials from stage 1 to stage 3, with those for stage 4 meant as an option.
The 93 students (3.53%) who scored 520 or more on the semester-initial TOEFL-ITP applied for exemption from the CALL course. Additionally, there were 39 dropout students (1.48%), who did not access the learning materials during a whole semester. Thus, the exempted students and dropout students were removed, and the remaining 2,499 students (94.99%) participated in this research.

The e-learning materials of the CALL course contained grammar, listening, and reading sections and included 493 units with a total of 751 quiz items. The quiz items were uploaded for students to study at the beginning of each semester and all of the learning activities occurred online.

Table 1: Course schedule in the spring semester 2016

<table>
<thead>
<tr>
<th>Stage</th>
<th>Deadline</th>
<th>Reading</th>
<th>Listening</th>
<th>Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Week 5</td>
<td>Reading1</td>
<td>Listening1</td>
<td>Grammar1</td>
</tr>
<tr>
<td>2</td>
<td>Week 10</td>
<td>Reading2</td>
<td>Listening2</td>
<td>Grammar2</td>
</tr>
<tr>
<td>3</td>
<td>Week 15</td>
<td>Reading3</td>
<td>Listening3</td>
<td>Grammar3</td>
</tr>
<tr>
<td>4(optional)</td>
<td>Week 21</td>
<td>Reading4</td>
<td>Listening4</td>
<td>Grammar4</td>
</tr>
</tbody>
</table>

3.2 Data Source

The data used in this research were collected from the CALL course server. The moment students practiced quiz items online, the learning behaviors were recorded in server logs concurrently. There were three types of learning logs in the CALL course server, including access to learning materials (access logs), completed quiz items (completion logs), and quiz answers (answer logs). A total of 7,413,397 learning logs were retrieved and analyzed from the server of the CALL course with 1,792,277 access logs, 1,117,375 completion logs, and 4,503,745 answer logs. The details of learning sessions were stored in logs, including user identifier, learning material id, quiz item id, access start time, access end time, completion flag, completion time, right answer id, and selected answer id.

3.3 Data Preparation and Analysis Procedures

Figure 1 illustrates the data preparation, analysis, and feedback procedures on the analysis server.

First, Search Query Language (SQL) queries were conducted to retrieve log data from the CALL course server and then log data were saved on SQL Server 2012-a database management system.

Second, the feature extraction phase was performed using reduced log files, which were cleaned by removing all useless, irregular, and incomplete data from the original CALL course logs. For example, sometimes the end time was not recorded when a student accidentally closed the web browser, or a student might do nothing for a long time with the website left open. In these cases, the related raw data were removed from the log files to reflect only normal learning activities of the students. Four primary variables were extracted through calculating or accumulating reduced log files: (1) Daily access items. (2) Daily access time. (3) Daily completion items. (4) Daily answer lists.

Third, the primary variables stored in the target database were analyzed and interpreted. The analysis phase included two sub-phases: (1) Descriptive analysis: an overview on the dataset was generated by summarizing and classification. The ratios of task completion at four stages were
summarized from the primary variable of daily completion items. Further, the active days of the students on the course were cumulated from the primary variable of daily access items, then the students were classified by the active days. (2) Correlation analysis: The relationship between the total time spent on the course and quiz scores was investigated from the primary variables of daily access time and daily answer lists.

Finally, based on the analysis results of massive log data, learning dashboard was provided. The learning dashboard focused on the learning processes and tendencies in the CALL course.

4. Results

4.1 Overview of Data Analysis

Table 2 summarizes the ratio of task completion at four stages (N=2499). The ratios of tasks completed at stage 1, stage 2, and stage 3 were 98.08%, 93.20%, and 88.52%, respectively. In contrast, the number at stage 4 significantly declined to 42.70%, since the learning materials of stage 4 were not required but optional. The results reveal that most of the students completed tasks at the required stages but 11.48% (n=287) of the students were still not completed after the third required stage. Moreover, 57.30% (n=1432) of the students stopped learning the materials at the optional stage. The number of students who dropped out at each stage increased continuously, thus the students who dropped out might need more support.

Table 2: Course stages and the ratios of task completion

<table>
<thead>
<tr>
<th>Stage</th>
<th>Total</th>
<th>Task completed</th>
<th>Not completed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>1</td>
<td>2499</td>
<td>2451</td>
<td>98.08%</td>
</tr>
<tr>
<td>2</td>
<td>2499</td>
<td>2329</td>
<td>93.20%</td>
</tr>
<tr>
<td>3</td>
<td>2499</td>
<td>2212</td>
<td>88.52%</td>
</tr>
<tr>
<td>4 (optional)</td>
<td>2499</td>
<td>1067</td>
<td>42.70%</td>
</tr>
</tbody>
</table>

The active days of each student were also investigated. One active day was defined as a day on which one student accessed learning materials. About half of the students (50.82%) accessed learning materials with 11 to 20 active days, and most of the students (90.88%) accessed learning materials with 30 active days or less. Since the duration of the CALL course was 21 weeks, most of the students (90.88%) performed self-regulated learning with a low degree (fewer than 22% of available days) of the learning activeness. Thus, the most of students might need more support to participate in the course.

![Figure 2. Scatters of total time spent on the course and quiz scores](image)
To investigate the quality of the learning outcomes, a scatter plot is generated. Figure 2 illustrates the relationship between the total time spent on the course and quiz scores. The results reveal that the students in the area A spent less time than others on the course but still obtained high scores. In contrast, the students in the area B spent above-average time on the course but scored less than 40, thus they might need more attention and support.

4.2 Learning Dashboard

From the results of behavioral analyses, several learning patterns of the students were identified. It was clearly necessary to pay special attention to the students who need guidance. Therefore, the learning dashboard for students and instructors was developed.

The learning dashboard is shown in Figure 3, which contains two parts: learning progress and behavioral tendencies. (1) The left part is a line chart to show the learning progress. One curve represents one student and the green curve means the class average of learning progress. The horizontal axis shows the dates and the vertical axis displays the counts of completed quiz units. (2) The right part is a parallel coordinate to present the overall tendencies of learning behaviors. The vertical axes show the quiz scores, the number of completed units, the active days, and the total time spent, respectively. One polyline connecting the vertices on the vertical axes represents one student. Meanwhile, the quiz scores are mapped to a six-color ribbon. When the quiz score is higher, the color of the polyline will be mapped from green to dark red.

For students, personal learning progress and personal behavioral tendencies were shown on the dashboard. They could support self-monitoring in a visual way by showing learning process and result. In addition, comparing personal behaviors to that of class average, the student easily found the position of self, leading to increased motivation and participation.

For instructors, learning progress in one class and behavioral tendencies in one class were shown. The constructors can see the student id and name by moving the mouse pointer to one line, and therefore it is easy to identify the students’ learning patterns and locate the student who need guidance.

5. Conclusion and Further Work

In this research, a learning support system for the CALL course in higher education was proposed. A total of 7,413,397 learning logs collected from 2,499 students' behaviors were analyzed. The results revealed that 11.48% of the freshman students still drop out during the SRL processes even on the mandatory courses in CALL, and most of the freshman students (90.88%) performed SRL with a low
degree (fewer than 22% of available days) of the learning activeness. The proposed system was based on actual learning logs, and it focused on the learning processes and tendencies. The system was provided for not only students but also instructors through the learning dashboard. On the one hand, students could conduct self-monitoring and reflect their behaviors in a visual way. On the other hand, instructors could identify learning behavioral patterns and grasp individual learning situation to provide one-on-one instructions. Moreover, the learning dashboards of the learning support system were easy to use on a variety of web browsers, since they were developed with JavaScript, which was a major browser scripting language.

For future work, the evaluation of usefulness and impact of the learning support system will be conducted. Further, in order to provide feedback more efficiently and identify at-risk students sooner, the learner models on learning progress and behavioral tendencies will be constructed.

Acknowledgements

The research has been partly supported by “Research and Development on Fundamental and Utilization Technologies for Social Big Data” (178A03), the Commissioned Research of the National Institute of Information and Communications Technology (NICT), Japan; and Grant-in-Aid for Scientific Research (S) No. 16H06304; Grant-in-Aid for Scientific Research (B) No. 25282059; Grant-in-Aid for Challenging Exploratory Research No. 26560122 from MEXT, Japan.

References

An Approach to Accent Visualisation for the Reduction of Vowel Pronunciation Errors

Tom A. F. ANDERSON, Barry Lee REYNOLDS & David M. W. POWERS

Abstract: In this paper, we introduce a novel mechanism for pronunciation evaluation for use in a computer-based application for the reduction of vowel errors. Unlike modern pronunciation feedback, which relies on the charts used by linguists or simply highlighting of incorrect segments, our two-dimensional maps situate the utterances produced by the learner in the context of the phonemes of the target accent.

Keywords: Pronunciation, phonemes, technology-enhanced language learning

1. Introduction

The concept behind this study is towards enhancing the activity of pronunciation practice. Learners know that they should devote time to pronunciation practice; however, efforts to substantially improve pronunciation require not only many sessions of practice but also integrated feedback. Learners need to understand how their pronunciation is improving, and the speaking quality needs to be reflexive. This means that learners should be able to practice the same thing many times with feedback that encourages the learner to alter their behaviour. Such practice induces the frequency effect, whereby “the greater the practice, the greater the performance” (Ellis, 2012, p.7).

A human teacher can provide targeted feedback to a learner, but there would be many advantages to computer-based training, including cost and repeatability (Chan et al., 2006). In mobile apps for language learning, the aspect of pronunciation is often restricted to reading (both alphabetic and phonetic) and listening. Reading and listening are necessary components for language learning; however, allowing users the opportunity to produce speech and providing visualisations that aid their understanding of their pronunciation errors may enhance the model of the student’s current communicative abilities.

2. Literature Review

2.1 Pronunciation

The typical feedback for pronunciation provided by language learning software has room for improvement. A common mechanism, adopted in software such as Rosetta Stone, a leading language learning software, is providing spectrographs, voice contours for visual inspection of the voice patterns (Witt, 2012), but learners are not experienced linguists, so may miss the point of what such representations provide. Feedback based on automatic speech recognition provides an indication of which words were not pronounced correctly, so this gives learners a better idea of what to focus on.

Most state-of-the-art pronunciation modules use automatic speech recognition with posit (ASR) (Golonka et al., 2012). Typically, these systems, such as EnglishCentral or Spexx, identify those sections of utterances that are less than ideal (see Witt, 2012 for a more comprehensive review of these programs and more). However, due to the nature of current pronunciation training systems, they often do not provide an indication of how the sounds of the language relate to each other, or how the current utterance may relate to other similar utterances. With an aim to improve computer-based pronunciation...
instruction, in this paper we present a biofeedback method that helps the learner to visualise their accent and how it fits into the context of the phonemic inventory of the language they are learning.

In contrast to systems that do not embed phonemes in their context, we have developed a neural network approach for visualising pronunciation. A Kohonen neural phonetic typewriter (KNPT) learns to represent the sounds of speech on a two-dimensional map with similar sounds located near one another (Kohonen, 1988; Kohonen, 2013). The underlying principle of the self-organising map is that information is arranged topographically, such that neurons that are near one another represent similar information. This allows viewing phonemes in relation to one another, and if necessary, by using an ensemble of maps, narrower contexts can also be visualised.

A two-dimensional map display allows the learner to get comfortable with a new representation of their voice. Although training a KNPT system can take a significant amount of time, once the values have been learned, it is very quick to classify the different sounds in a stream of speech. The learner can say something and the display will show the current state of their voice, along with the trajectory of the recent path of their voice. As neurons corresponding to similar inputs are located on nearby regions, a trajectory will generally pass from one zone to another through a transition zone. By tuning the maps, representations of voices and accents in the context of an individual speaker or an accent group are generated.

2.2 Form Focused Instruction

In recent years, isolated form-focused instruction has received renewed interest and a welcome response by second language acquisition researchers (Spada & Lightbown, 2008). Isolated form-focused instruction for pronunciation errors could be helpful in deterring first language interference with second language pronunciation. In terms of the kinds of mistakes that occur in pronunciation, Witt (2012) differentiated between phonemic and prosodic errors, the former being our current focus. Phonemic mistakes may arise in two forms: (1) severe - a phoneme is replaced by another, omitted, or an extra phoneme is produced; or (2) accented - a phoneme is pronounced with an accent. Its sound is thus different than a native speaker would produce. Both types of errors may affect the intelligibility of the learner.

Usually during form-focused instruction on such pronunciation errors, the classroom teacher provides oral corrective feedback in the form of recasts (implicit feedback) or metalinguistic explanation (explicit feedback). Although both are frequent occurrences in the second language classroom, explicit feedback has been shown to have a greater effect on language acquisition (Ellis, Loewen, & Erlam, 2006). However, in the limited class time a teacher has with students, all pupils cannot be given corrective feedback on their oral language production.

The effects arising from limited time for pronunciation feedback are sometimes alleviated by pairing language learners with tutors outside the classroom, which can result in not only the learners perceiving themselves as having improved but also showing said improvement on formal assessments (Lynch & Maclean, 2003). Still, this is not always a practical option since it cannot be guaranteed that a more capable peer or tutor can be secured for every language learner.

Form-focused explicit feedback given by the computer is a probable solution to this conundrum. In addition, one of the benefits of oral corrective feedback provided by the computer is the reduction in the potential of affective damage that can occur when language learners receive feedback from a teacher in front of their classroom peers. Language learners have an emotional response to the feedback delivered by teachers and when this response induces anxiety, the potential for negative effects on language learning increases (Agudo & de Dios, 2013). The computer can make accessible the type of feedback needed by all language learners within a comfortable context and environment.

3. Overview

A mobile application is used for accent reduction. A system flowchart of the system is depicted in Figure 1. For directed learning, a prompt is selected for a learner based on the system learner model. Note that learners may be involved in selecting the order of prompts, or produce spontaneous speech.
Speech is evaluated in the context of the accent of the speaker, with a goal to provide feedback quickly, within 300 milliseconds. The learner model is updated as interaction increases with the system.

3.1 Improving the Pronunciation of Vowels.

As vowels are produced by an uninterrupted outflow of air, the sounds of vowels appear on the maps in continuous trajectories (averaged as traces with thickness reflecting variance). There are multiple ways to get similar sounds. However, the trace of a target accent for the pronunciation of the vowels is narrower than the trace for the corresponding foreign accented pronunciation. In other words, the sounds of the vowels that the foreign speakers produce are naturally and consistently out of the target zone. A speaker’s first task is thus to pronounce the vowel closer to the target. The pronunciation of steady-state vowels can be changed and shaped by moving various articulators. The idea is to get the speaker to produce vowel sounds in isolation that are more similar to the target pronunciation. Next, speakers should produce the vowel sounds accurately in the context of isolated words.

In continuous speech, when many words are strung together, the pronunciation of each particular sound is much less important. In contrast, when a single word is pronounced in isolation, each word and each of its constituent phonemes is expected to be pronounced clearly and appropriately. The eventual goal is to aid the speaker to produce speech that is more intelligible or less accented. It is not guaranteed that reduction in accentedness in the pronunciation of isolated words will cross over to the regular speech patterns. However, during the process of learning how to make the sounds of an individual word be closer to the target pronunciation, the learner will gain an understanding of how to position their articulators to produce certain sounds that were previously less familiar to them. The next step would be to help learners to understand how the sounds of their own voice, in continuous speech, can be shaped to produce speech that is more like the target, and to give them an ability to practice shaping spoken words with less of an accent.

3.2 Speech Representation

Ensembles of self-organizing maps (SOMs) were trained on the voices of native speakers (general Australia, educated Melbourne) and a target group (Chinese background) using data from the AusTalk corpus (Burnham et al., 2011; Burnham et al., 2009; Wagner et al., 2010). The maps provide a visual representation of the speech of the learner in phonemic context. The speech of the learner is pre-processed into 39-feature mel-frequency cepstral coefficients (MFCCs), commonly used in automatic speech recognition. The MFCCs are then compared to codebook vectors. Although the initial training of the system must be performed offline, the evaluation and updates to the learner model can be performed near real-time. For more details, see Anderson and Powers (2016).

3.3 Implementation

Client – A microphone is used to obtain audio input. This may be the device microphone of the mobile computing device (often an array microphone in modern phones and laptops), but an external microphone may result in better sound quality.
System – As the learner speaks, their voice is analysed using the system and feedback is presented. Speech is shown on a map using a selection of pronunciation samples, as in Figure 2. Learners interactively explore the differences between how their speech and pre-recorded utterances are rendered, thereby improving the understanding of their speech in the context of the speech of others.

Gamification – Games can increase motivation. In a meta-analysis on digital game based learning for English as a foreign language, Chui et al. (2011) found that language learning games positively affect learning but meaningful and engaging games can yield a greater positive effect than drill-and-practice games. When learners are immersed in digital game based learning environments for more than a month, they gain automisation of language knowledge (Kao, 2014). We feel that automisation is a necessary aspect for improving pronunciation. In terms of future directions for our research, we aim to apply our SOM-based phoneme visualisation to game-based learning. The difficulty level can be increased by chaining multiple utterances together, providing fewer hints, or by narrowing the tolerance, which requires a learner to produce speech closer to the target.

4. Conclusion

In this paper, a mobile learning environment for pronunciation was proposed. Learners receive real time feedback about their speech and achieve goals in a video game by modifying their speech. Learners are presented with lessons sequenced so that they may gradually improve toward the target. In future research, the usability of the system for accent reduction would be evaluated, along with additional aspects for promoting learning such as gamification.
Acknowledgements

The AusTalk corpus was collected as part of the Big ASC project (Burnham et al. 2009; Wagner et al. 2010; Burnham et al. 2011), funded by the Australian Research Council (LE100100211). See: https://austalk.edu.au/ for details

References


Chinese Grammatical Error Detection Using a CNN-LSTM Model

Lung-Hao LEEa, Bo-Lin LINb,c, Liang-Chih YUb,c & Yuen-Hsien TSENGa*

aGraduate Institute of Library and Information Studies, National Taiwan Normal University, Taiwan
bDepartment of Information Management, Yuan Ze University, Taiwan
cInnovation Center for Big Data and Digital Convergence, Yuan Ze University, Taiwan
*samtseng@ntnu.edu.tw

Abstract: In this paper, we proposed a Convolution Neural Network with Long Short-Term Memory (CNN-LSTM) model for Chinese grammatical error detection. The TOCFL learner corpus is adopted to measure the system performance of indicating whether a sentence contains errors or not. Our model performs better than other neural network based methods in terms of accuracy for identifying an erroneous sentence written by Chinese language learners.

Keywords: Grammatical error diagnosis, deep neural networks, Chinese as a foreign language

1. Introduction

Chinese as foreign language learners usually make different kinds of grammatical errors during second language acquisition process (Lee et al., 2016a). Automated grammatical error detection and correction are emerging important research directions and a number of competitions have been organized to encourage innovation (Leacock et al., 2014). Recently, the Natural Language Processing Techniques for Educational Applications (NLPTEA) workshops have hosted a series of shared tasks for Chinese grammatical error diagnosis (Yu et al., 2014; Lee et al., 2015; Lee et al., 2016b). All of these activities attracted global participations and enhanced research developments.

Language models have been adopted to detect various types of Chinese errors written by US learners (Wu et al., 2010). A probabilistic inductive learning algorithm has been proposed to diagnose Chinese grammatical errors (Chang et al., 2012). Linguistic rules have been manually constructed to detect Chinese erroneous sentences (Lee et al., 2013). Support Vector Machine based classifiers have been used to explore useful features for detecting word-ordering errors in Chinese sentences (Yu and Chen, 2012). A sentence judgment system has been developed to detect grammatical errors in Chinese sentences using both n-gram statistical analysis and rule-based linguistic analysis (Lee et al., 2014).

Gated recurrent neural network models have been explored to select the best prepositions for Chinese grammatical error diagnosis (Huang et al., 2016). In recent NLPTEA workshops (Lee et al., 2015; Lee et al., 2016b), neural approaches have been explored for identifying Chinese grammatical errors. This observation motivates us to explore neural networks to detect errors written by Chinese learners.

This study describes our proposed Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) model, a kind of deep neural network, for Chinese grammatical error detection. The TOCFL learner corpus is used to evaluate and compare performance. Error detection systems that indicate grammatical errors in a given sentence are useful to learners for computer-assisted language learning.

2. Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM)

Figure 1 shows our Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) architecture for Chinese grammatical error detection. An input sentence is represented as a sequence of words. Each word refers to a row looked up in a word embedding matrix generating from Word2Vec (Mikolov et al., 2013). A single convolution layer is
adopted. We use convolutions over the sentence matrix to extract features. The full convolutions are obtained by sliding the filters over the whole matrix. Each filter performs the convolution operations on the sentence matrix and generates a feature map. A pooling layer is then used to subsample features over each map. We apply the max operation to reduce the dimensionality for keeping the most salient features. To capture long-distance dependency across features, LSTM is used in the sequential layer for vector composition. After the LSTM memory cells sequentially traverse through all feature vectors, the last state of the sequential layer is regarded as input for neural computing. The final softmax layer then receives computing results and uses it to classify the sentence.

During the training phase, if a sentence contains at least one grammatical error judged by a human, its class is labeled as 1 and 0 otherwise. All the sentences with their labeled classes are used to train our CNN-LSTM model to automatically learn all the corresponding parameters in this model.

To classify a sentence during the testing phase, the sentence goes through the CNN-LSTM architecture to yield a value corresponding to the error probability. If the probability of a sentence with class 1 (i.e., with errors) exceeds a predefined threshold, it is considered as true as an erroneous sentence and false otherwise.

Figure 1. The illustration of our CNN-LSTM model for Chinese grammatical error detection.

3. Experiments and Evaluation Results

The experimental data came from the TOCFL learner corpus (Lee et al., 2016a), including grammatical error annotation of 2,837 essays written by Chinese language learners originating from 46 different mother-tongue languages. Each sentence in each essay is manually labeled. The result is that a total of 25,277 sentences contain at least one grammatical error, while the remaining 68,982 sentences are grammatically correct (an unbalanced distribution with 26.82% sentences having grammatical errors). Five-fold cross validation evaluation was used to measure the performance.

To implement the system, a python library Theano was used. For Word2Vec representation, Chinese Wikipedia 2016 was trained to generate 300 dimensional vectors for 655,247 words and phrases. The number of filters was 300 and their length is 3. The number of iteration (i.e., epochs) was set up as 5 to learn the CNN-LSTM network parameters. If the error probability of an input sentence exceeds 0.3, it was considered as an erroneous sentence.

The following three methods were compared to demonstrate their performance. (1) CNN only: this method only considers the CNN part of our proposed model. (2) LSTM only: this approach only focuses on the LSTM part of our proposed model (3) CNN-LSTM: this is our proposed model for Chinese grammatical error detection.

Table 1 shows the results. The CNN only and CNN-LSTM model respectively had the best recall and precision. Considering the tradeoff, the LSTM only model reflected the best F1-score of 0.4859 (the improvement compared to the lowest F1-score is 5.4%). In addition to best precision, our
The proposed CNN-LSTM model also achieved the best accuracy of 0.6905 (the improvement compared to the lowest accuracy is 12.77%).

Table 1: Evaluation on Chinese grammatical error detection.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN only</td>
<td>0.6123</td>
<td>0.3745</td>
<td>0.6488</td>
<td>0.4717</td>
</tr>
<tr>
<td>LSTM only</td>
<td>0.6599</td>
<td>0.4179</td>
<td>0.6049</td>
<td>0.4859</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>0.6905</td>
<td>0.4439</td>
<td>0.5057</td>
<td>0.4610</td>
</tr>
</tbody>
</table>

4. Conclusions

This study describes the CNN-LSTM model for Chinese grammatical error detection. We use the TOCFL learner corpus to demonstrate system performance. Our system achieved the best accuracy of 0.6905 for predicting whether a given sentence contains grammatical errors or not, which roughly corresponds to 7 out of 10 input sentences were judged correctly under the unbalanced error distribution.

Acknowledgements

This study was partially supported by the Ministry of Science and Technology, under the grant MOST 103-2221-E-003-013-MY3, MOST 105-2221-E-155-059-MY2, MOST 106-2221-E-003-030-MY2 and the “Aim for the Top University Project” and “Center of Language Technology for Chinese” of National Taiwan Normal University, sponsored by the Ministry of Education, Taiwan, ROC.

References


Global Collaborative Learning Support System for Facilitator Collaboration: First Phase Development Report

Yoshiko GODA\textsuperscript{a*}, Masanori YAMADA\textsuperscript{b}, Yumi ISHIGE\textsuperscript{c} & Junko HANDA\textsuperscript{d}

\textsuperscript{a}Research Center for Instructional Systems, Kumamoto University, Japan
\textsuperscript{b}Faculty of Arts and Science, Kyushu University, Japan
\textsuperscript{c}Faculty of Modern Social Studies, Otemae University, Japan
\textsuperscript{d}Institute for Service Innovation Studies of Meiji University, Japan
*ygoda@kumamoto-u.ac.jp

Abstract: This research reports the initial design and development of a global collaborative learning (GLoCL) support system for collaborative work among multiple facilitators. Among the system’s main features are (1) support for facilitators collaborating on the design and implementation of global project-based learning; (2) improved feedback and enhanced quality of student interaction; and (3) reduction of the administrative burden on facilitators. The system will be developed in two phases. The poster presentation introduces the system’s overall design and describes the first phase of system development.

Keywords: global learning, facilitation, CSCL, community of inquiry

1. Introduction

In Japan, internationalization and global human development in higher education are strongly encouraged (Central Council for Education, 2010). The cultivation of global human resources requires an approach that emphasizes learning through authentic activities and measurement of acquired knowledge in terms of learning outcomes. By eliminating restrictions related to time and space, Information Communication Technology (ICT) can provide authentic global learning opportunities for students. However, only a few examples of ICT-based global collaborative learning have been reported in Japan. One survey (Goda et al., 2014) reported that only 2 out of 327 university students had learned in collaboration with foreign students online. They identified time differences, curriculum and school timetables, legal issues, infrastructures, and teachers’ altered role as facilitators of collaborative learning as obstacles to implementing global collaborative learning. As computer-supported collaborative learning (CSCL) gains in popularity, tools and systems have been developed in support of CSCL (e.g., Knowledge Forum, WorldWatcher), but most of these center on face-to-face classroom use and take no account of global learning settings and collaboration among multiple facilitators.

The ultimate goal of our research project is to design and develop a support system for global collaborative learning for both students as active learners and teachers as facilitators. The purpose of this research is to report on the initial design and development of a global collaborative learning support (GLoCL) system that supports collaborative work by reducing the burden on facilitators and making it easier to design and implement global project-based learning online.

2. Overall Design and Two Phases of System Development

The overall design of the proposed GLoCL system consists of ten functions. Five functions are common to students and teachers (Table 1), and five provide facilitator support (Table 2). Effective CSCL design based on community of inquiry (CoI, Garrison, Garrison, Anderson, & Archer. 2000) were introduced for English learning (Goda and Yamada, 2012). CoI was employed as a framework to design the system for quality interactions among students. In the first of two phases of system development, functions (other than Top Page and Submission) have been developed as a Moodle module. Three of the system’s
key features are (1) support for facilitators collaborating on the design and implementation of global project-based learning; (2) improved feedback and enhanced quality of student interaction; and (3) reduced administrative burden. Each of the system’s functions was designed to deliver more than one of these features. In the case of (2), for example, the facilitation function visualizes the social presence and cognitive presence of a community of inquiry. Ongoing discussion in each group is supported by the application of text mining using tf-idf, enabling facilitators to identify need-help groups (Yamada, et al. 2016). Where feedback is needed during group discussion, the Facilitation function provides easy-to-use comment templates for effective feedback within each learning process. The Group Management function uses questionnaire results for manual or automatic generation of homogeneous or heterogeneous groups in relation to a specific trait or characteristic.

<table>
<thead>
<tr>
<th>Function</th>
<th>First Phase</th>
<th>Second Phase</th>
<th>System Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Student</td>
<td>Teacher</td>
<td>Student</td>
</tr>
<tr>
<td>Top Page</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Check reminders and what’s new after logging out last time; see world clock; get reminders for upcoming tasks; access information on cultural differences.</td>
</tr>
<tr>
<td>Chat</td>
<td>Chat in one’s group. Post, edit, and delete comments. Reply to others’ comments. Use emoticons. Attach a file.</td>
<td>See a list of groups. Check and monitor chat in each group. Make comments on group chat.</td>
<td>Revise when necessary</td>
</tr>
<tr>
<td>Task Scheduling</td>
<td>Plan schedules. Share tasks and decide their dues. Assign tasks. Check tasks in Calendar.</td>
<td>See a list of groups. Check and monitor the task assignments and schedule of each group.</td>
<td>Revise when necessary</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Take questionnaire</td>
<td>Preview questionnaire</td>
<td>Revise when necessary</td>
</tr>
<tr>
<td>Submission</td>
<td>Not yet available</td>
<td>Not yet available</td>
<td>Submit tasks as a group or individually</td>
</tr>
</tbody>
</table>
Table 2: Design of Global Collaborative Learning Support System: Facilitator Support Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>First Phase</th>
<th>Second Phase</th>
<th>System Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire</td>
<td>Manage questionnaires.</td>
<td>Revise when necessary</td>
<td>(3) Questionnaire is used for grouping. Appropriate and adaptive questionnaire can be used for grouping.</td>
</tr>
<tr>
<td>Manager</td>
<td>Manage groups. Get the questionnaire results and make groups based on the results (homogeneous group vs. heterogeneous group).</td>
<td>Revise when necessary</td>
<td>(3) Manual or automatic grouping with questionnaire results available.</td>
</tr>
<tr>
<td>Group Manager</td>
<td>Log (Chat, Questionnaire, Behavior) download.</td>
<td>Revise when necessary</td>
<td>(3) The log is downloadable for grading or research purposes.</td>
</tr>
<tr>
<td>Facilitation</td>
<td>Monitor students’ process of learning. Check the visualized SP and CP in a group and of each student. Access the group chat with some problems and give feedback and support with comment support function.</td>
<td>Check “Facilitator Guideline,” final “Project Design,” Cultural Feature Information</td>
<td>(2) Visualization of social presence and cognitive presence allow facilitators to find salient features of the groups in discussion. Comment template along with SP and CP process help facilitators provide more appropriate feedback to individuals and the group.</td>
</tr>
<tr>
<td>Project Design</td>
<td>Share current course or project information among facilitators. Design a project for students.</td>
<td>Revise when necessary</td>
<td>(1) Design project based learning for global learning considering purposes and situations of all participant facilitators.</td>
</tr>
</tbody>
</table>

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26284079.

References


A Telexistence Robot Combined with Virtual Reality for Teaching English

Wen-Chi Vivian WU*, Tosh Yamamotob, Cheng-Hao Hu& Rong-Jyue Wang*

aDepartment of Foreign Languages and Literature, Asia University, Taiwan
bThe Center for Teaching and Learning, Kansai University, Japan
cDepartment of Electronic Engineering, National Formosa University, Taiwan
*orffwang@nfu.edu.tw

Abstract: Learning English has been a global trend since English is widely regarded as an international language for effective communication in the global village where we live today. However, Taiwan is a non-English speaking country; the lack of professional English teachers in Taiwan is still a serious problem both in the countryside and the city. In order to solve this problem, in our study the telexistence robot comes to play. To make teachers feel a sense of telepresence, more feedbacks were prepared for the scene and more focal points for the scene were enriched on teaching scenario. The robot was designed in the newest technological way where virtual reality (VR) and diverse sensors were employed to receive telepresence and feedback. Teachers could become even more engaged in teaching scenarios and operate the robot from a distance through wireless communication technology. The application of telexistence robot has been commonly recognized, and users can operate the robot by sensors whose wireless signal transmits from the user to the robot. The user wears wearable devices on both arms, both feet and the head to control the remote robot to display the same body movement and position. In this way, the application of the telexistence robot to teaching in the elementary school seems practical and feasible. The findings of this study reveal that RALL (Robot Assisted Language Learning) group did make students more motivated and create a more joyful learning environment, compared to the traditional teaching method without RALL.

Keywords: English Learning, RALL, VR, telexistence robot, guidelines, scenario

1. Introduction

In the last few years, it has become more and more popular that using robots for supporting teaching and learning (Hung et al., 2012). Through the TPR (Total Physical Response) teaching method can really make learners reduce a sense of nervousness and learn a language with no pressure (Asher, 1982). However, how to design such a robot that children love on the one hand and how to design the instructional program with above teaching methods on the other are extremely important issues that we care about. Goetz et al. (2003) claim that the humanlike robot is more suitable than the machine-like robot to be an assistant or a teacher in the class. Tanaka, et. al. (2014) show that it has more advantages that teaching students through a remote robot is much more prominent than the traditional teaching method using TV and Video that only provide children with sound and video with one way communication, and thus children are unable to discuss problems immediately with teachers who are on the TV or Video screen. That is why we create the telexistence robot to improve the learning environment. Kwon, et. al. (2010) show that using a telexistence robot to teach students who live in the countryside resolved the trouble that many teachers are unwilling to travel and teach there. Thus, in the paper, the telexistence robot is used as a teacher in the class and a human teacher can operate the robot from a distance with wearing the wearable devices and the virtual glasses over the wireless communication.

2. Hardware and Software Design of the Telexistence Robot

The design of the telexistence robot in the paper should become like an anthropomorphic robot or humanlike robot with the size of 120 cm in height and 40 cm in width and it should look adorable, cute,
kindly and friendly. Further, the body structure is composed of aluminum bars to connect with each other with screws, and the AI-motors serve the movement of the components. For example, one hand has five motors to demonstrate its movements. Two motors serve the movement of the head. The signals from the robot are sent from the controller or a transmitter board connected to a gyro sensor, a bending sensor, and a pressure sensor that are installed in the body of the robot. And the signals to the robot are sent to the controller or a receiver board connected to the motors installed in the various body parts of the robot. When a teacher wears wearable devices and the VR glasses, he/she can feel as if being in the remote class to interact with students there.

The signal delivery system and the motion detection of the different parts are as follows: In order to make students improve their English before and after class, researchers particularly designed an English learning system that collocates with the kinematic motions or body language and the teaching method based on the CLT and TPR. The robotic kinematics are computer programmed precisely to make the body movements smooth and natural. In this way, the robot can stimulate and arouse students’ learning interest and desire.

Moreover, the sound recognition system is incorporated with the robotic system with the CLT, TRP and body language. Furthermore, MEILA (Mobile English Idioms Learning Assistant) system is installed to execute lessons before and after class through the interactive performance with the robot.

![Wireless Signal Delivery System](image1.png)

**Figure 1. Wireless Signal Delivery System**

![Scenes: the Demonstration of Motions](image2.png)

**Figure 2. Scenes: the Demonstration of Motions**

### 3. Method and Outcome

The experiment of the study was conducted with two groups. They are RALL Group (with the robot) and Conventional Group (without the robot). Both groups were given the same amount of the learning for the experiment. The learning period of the whole study lasted five periods, each of which was 50 minutes in length. In the lessons, students learned how to determine the meaning of sentences with rising tone or falling tone. From the academic standpoint, students learning English through r-learning and TPR with the telexistence robot had better performance than the students learning English without robot. The students in the RALL group enjoyed the class well and felt interesting in taking classes. In addition, the Conventional Group instructed with a real teacher, where the instructional program was
based on the teaching method of r-learning and TRP, including the intonation patterns of three types of questions, story-telling and sentence guessing.

Figure 3. Demonstration of the Telexistence Robot in the Remote Class

It follows that both groups demonstrated significant improvement. Following the experiment, we also interviewed students individually to analyze their thoughts about taking classes with instruction by a robot and by a human teacher. For RALL Group, some students said ‘never thought the robot could be a teacher and teach us English’. For the Conventional Group, some said that ‘the robot was not attractive enough compared with a real teacher’. Above all, students thought that they were motivated to learn more and the experiment gave them positive learning experience.

4. Conclusion

In the paper, our aim is to solve the problem that is the lack of professional English teachers in the countryside in Taiwan and that most of teachers are unwilling to travel and teach there. Therefore, we created the telexistence robot that can be controlled and operated from a long distance. The problem was resolved now to make the professional English teachers in the city be able to teach anywhere. Furthermore, such system will replace the traditional teaching systems that employ TV-learning, radio-learning and the like. Moreover, the teaching method of r-learning, TPR and CLT enhanced with the telexistence robot is expected to establish much more high performance and excellence in teaching in English. Last, but not the least, the result of the analysis of the personal interview revealed that the telexistence robot in the class demonstrated more learning efficiency than the class without it.

Acknowledgements

This research was partially supported by the Ministry of Science and Technology in Taiwan through Grant MOST 102-2221-E-150-059-MY3 and MOST 104-2511-S-126 -003 -MY2.

References


Integration of Peer Assessment and Shadowing Strategies for Improving the Oral Performance of EFL Learners

Siao-Cing GUO a & Ting-Chia HSU b *

a Department of Applied Foreign Languages, National Taipei University of Business, Taiwan
b Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan

*ckhsu@ntnu.edu.tw

Abstract: This study integrated the peer assessment and shadowing strategies in an English course. English is a mandatory subject throughout the education system in Taiwan. In this study, the average age of the participants was 17. For the experimental group, the teachers provided native-like English audio input from the Internet. The students linked to and listened to the audio content. They then made their own recordings and uploaded them to the cloud drive so that their peers could listen to them and evaluate their speaking performance based on the rubrics the teacher provided. For the control group, the same process was followed as with the experimental group but the students did not access the audio materials for shadowing. After four weeks, we compared the speaking performance of the two groups and found that the students in the experimental group outperformed those in the control group.

Keywords: peer assessment, oral speaking, English as a foreign language, shadowing

1. Introduction

English is a lingua franca in this globalized world and is an important means of gaining access to and information about the online world. English plays a prominent role in students’ academic achievement and their future careers. Because Taiwan is not an English-speaking country, learning the language poses a greater challenge for students. Students generally view English as a subject to study rather than as a language for communication, and they do not have the opportunity to hear or use the language in their daily lives. Thus, their exposure to and input of the language is insufficient.

Since students are in a mono-linguistic environment, their exposure to the language is limited. Most English courses focus on reading and writing. The source of the listening input is limited and the opportunity to speak the language is also scarce. Hence, English teachers need to provide sufficient input and design various activities to improve students’ language proficiency. In this present study, the researchers conducted a study which integrated peer-review in the oral practice with a shadowing strategy. The researchers attempted to investigate the effect of the shadowing strategy to see if it could aid in the development of the English language.

2. Literature Review

Language is a system that encompasses listening, speaking, reading, and writing. For non-native speakers of English, speaking is an important skill that students have to work on (Boonkit, 2010). Creating opportunities for students to practice speaking is one of the key components in English teaching. Although many English courses and course books separate speaking and listening skills in instruction, according to Tavil (2010), the two skills are as important as each other. Tavil also stressed the importance of the integration of listening and speaking in teaching. The more practice students have in listening and speaking, the more comfortable and confident they will become.
2.1 Shadowing

When it comes to speaking in a foreign language, it is perceived to be better to be close to the pronunciation of native speakers of the target language. Teachers usually utilize speaking tasks to let students imitate native speakers. One speaking strategy that is often used to practice speaking is called “shadowing” (Takeuchi, 2003). Shadowing is the action of following the speech one hears and of vocalizing the information as one listens (Tamai, 2001). Murphey (2001) identified different forms of shadowing tasks including complete, selective, and interactive. Complete shadowing refers to learners repeating everything they hear, whereas in selective shadowing, learners only select certain words or parts of the speech to repeat. In interactive shadowing, learners use selective shadowing but add questions or comments afterwards. Shadowing enhances the students’ listening and communication skills. The speaking source that the learners listen to scaffolds their knowledge of the target language form (Apple, 2006).

2.2 Peer Review

Working with others is important for developing social and cooperative skills which students are likely to need in their school lives and their future careers. Some students may find it difficult because they do not think they can find things to say or they believe that their language proficiency is insufficient (Singh, M.S, 2007). The interaction among their peers can increase students’ English ability and provide them with more opportunities for learning, in particular for less proficient learners (Genesee et al., 2005).

3. Methods

3.1 Participants

A total of 57 17-year old English major students were recruited to participate in this oral audio study. They were divided into the control group with 30 students and the experimental group with 27 students. The students each had a smartphone with wireless Internet access so that they could access the audio on the Internet, record their own speaking, and upload their recorded files to the cloud drive.

3.2 Measuring Tools & Procedure

Oral questions from the simulated General English Proficiency test were used in the study. The questions were selected and reviewed by three English teachers to ensure that they were suitable for testing the students’ English ability. The students’ oral skills were tested prior to the study. The results from the independent t-test of the pretest showed no statistically significant difference ($t=0.10, p>0.05$) in the oral skills of the students from the control group and the experimental group. The students’ speaking ability was therefore at a similar level.

The two groups of students were given two different oral assignments in which they had to respond to the oral questions and record their answers. After the answers were recorded, they had to upload them to the Internet platform for their peers to evaluate. The students from both groups followed the same procedure, except that the students from the experimental group had to listen to the sample videos and shadow the recording prior to giving their own responses to the questions. The students from the control group did not listen to the sample recording before they gave their responses.

4. Results & Conclusion

In order to compare the effect of the shadowing strategy on the peer-review speaking project, the $t$-test technique was used to analyze the speaking scores from the students’ two speaking assignments. The results of oral assignments 1 and 2 showed statistically significant differences between the control
group and the experimental group. The results from the two oral assignments can be found in Tables 1 & 2.

Table 1: The score comparison of oral assignment 1

<table>
<thead>
<tr>
<th>English test (Oral 1)</th>
<th>N</th>
<th>Means</th>
<th>SD</th>
<th>t</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>n=27</td>
<td>86.59</td>
<td>3.82</td>
<td>2.11*</td>
<td>55</td>
<td>0.03</td>
</tr>
<tr>
<td>Control</td>
<td>n=30</td>
<td>82.47</td>
<td>9.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.05

Table 2: The score comparison of oral assignment 2

<table>
<thead>
<tr>
<th>English test (Oral 2)</th>
<th>N</th>
<th>Means</th>
<th>SD</th>
<th>t</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>n=27</td>
<td>85.96</td>
<td>2.57</td>
<td>2.04*</td>
<td>55</td>
<td>0.04</td>
</tr>
<tr>
<td>Control</td>
<td>n=30</td>
<td>84.67</td>
<td>2.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of this study indicate that the speaking practice in the integration of the shadowing strategy improved the students’ oral ability. With regular speaking practice, students can improve their language. However, the use of the shadowing strategy allows learners to imitate the language that has a native-like quality and enables them to articulate better in their oral production.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology in Taiwan under contract numbers: MOST 105-2628-S-003-002-MY3.

References


Effect Analysis of Students’ Learning Styles on Learning Experience with Lecture Videos Played at Different Playback Speeds

Toru NAGAHAMA* & Yusuke MORITAb

*aGraduate School of Human Sciences, Waseda University, Japan
bFaculty of Human Sciences, Waseda University, Japan
*tnagahama@aoni.waseda.jp

Abstract: The goal of this study was to clarify effects of students’ learning styles on their learning experience with lecture videos played at different playback speeds. We focused on the interactions between students’ learning styles and video playback speed. In our experiment, participants’ learning styles were categorized by Felder’s Index of Learning Styles and 35 verbal students and 40 visual students learned about the network infrastructure with lecture videos played at original speed, 1.5× speed, and 2.0× speed. The comprehension test results indicated that the video playback speeds and the students’ learning styles did not influence the comprehension test scores. The subject evaluation results indicated that there were significant interactions for their learning experience. Consequently, the possibility that students’ learning experience could differ to their learning styles when they learn with hi-speed lecture videos was discussed.

Keywords: lecture video, online learning, high-speed video, variable-speed functionality

1. Introduction

As the popularity of MOOC (Massive Open Online Course) continues to grow all over the world, many studies have examined the importance of lecture videos. Guo, Kim, and Rubin (2014) studied lecture videos on MOOC by analyzing a dataset containing some 7 million instances of students watching lecture video. Guo et al. (2014) suggested that the number of course participants paying attention to a lecture video begins to decrease significantly when the video duration were longer than 6 minutes and instructor’s speaking rate was fairly slow. Nagahama and Morita (2017) studied the efficacy of using variable-speed playback functionality to watch lecture videos at high speeds and indicated that video playback speed: original speed; 1.5× speed; 2.0× speed did not influence the learning outcomes. Nagahama and Morita (2017) also suggested that watching lecture videos would highly increase cognitive loads. On the other hand, Felder’s index of learning styles (F-ILS: Felder and Henriques, 1995) has been used for researches which examine the relationship between learner characteristics and learning experience (Morita, Koen, Ma, Wu, & Jhendran, 2005; Oyama, Murakami, Taguchi, & Matsushita, 2010). The goal of this study was to clarify effects of students’ learning styles on their learning experience with lecture videos played at different playback speeds.

2. Methods

In our experiment, 35 verbal learners and 40 visual learners (all Japanese), whose learning styles were categorized by F-ILS, learned about the network infrastructure of a high school information science department with lecture videos. First, before watching the lecture videos, we gave a comprehension test (the pre-video test) to assess their pre-existing knowledge of the theme in the lecture video. Next, we divided 75 participants into three groups; (a) 1.0 group, who watched the lecture video at original speed; (b) 1.5 group, who watched the lecture video at 1.5× speed; (c) 2.0 group, who watched the lecture video at 2.0× speed. Next, each group of participants watched the lecture video. Next, after watching the
lecture video, the participants were given a post-video test. Finally, all participants were shown condensed versions of lecture video (lecture video digests) and were asked to complete a sheet of questions.

The lecture videos, the comprehension test, and the sheet of questions were the same as the ones that Nagahama and Morita (2017) used in their experiment. The comprehension test consisted of 20 problems, including 11 playback problems and 9 application problems. The questions consisted of two questions concerning comprehension, tow questions concerning speaking style, two questions concerning level of interest, three questions concerning concentration, three questions concerning ease of listening, three questions concerning ease of watching, four questions concerning whether students liked the speed and duration of the video, and five questions concerning whether students liked the video. The subjective opinions of the participants were surveyed using a five-point Likert scale.

3. Results

3.1 Comprehension Test

We determined the comprehension score and conducted a two-way ANOVA using the students’ learning styles as the first factor (F-ILS factor) and the video playback speeds as the second factor (speed factor).

For the playback problems, the ANOVA result indicated no significant interaction, \(F(2, 69) = 0.65, p > .05\). An analysis of main effects indicated no significant difference for the playback speed factor, \(F(2, 69) = 0.53, p > .05\), and for the F-ILS factor, \(F(1, 69) = 0.82, p > .05\). For the application problems, the ANOVA result indicated no significant interaction, \(F(2, 69) = 0.59, p > .05\). An analysis of main effects indicated no significant difference for the playback speed factor, \(F(2, 69) = 0.42, p > .05\), and for the F-ILS factor, \(F(1, 69) = 0.58, p > .05\).

3.2 Subjective Evaluations

We computed the mean scores for each question and conducted a two-way mixed ANOVA using the students’ learning styles as the first factor (F-ILS factor) and the video playback speeds as the second factor (speed factor). Significant interactions were found in four out of 24 questions. We focused on these four interactions in this paper. Table 2 shows the mean scores with the ANOVA results.

<table>
<thead>
<tr>
<th>Question</th>
<th>Verbal</th>
<th>Visual</th>
<th>F-Value</th>
<th>Speed</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4.</td>
<td>3.7</td>
<td>4.3</td>
<td>2.9</td>
<td>3.8</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
<td>(0.89)</td>
<td>(1.26)</td>
<td>(1.01)</td>
<td>(0.88)</td>
</tr>
<tr>
<td>Q11.</td>
<td>1.9</td>
<td>2.1</td>
<td>3.7</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(0.96)</td>
<td>(1.12)</td>
<td>(1.08)</td>
<td>(1.17)</td>
</tr>
<tr>
<td>Q15.</td>
<td>4.3</td>
<td>4.2</td>
<td>3.5</td>
<td>4.2</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>(0.80)</td>
<td>(0.61)</td>
<td>(1.20)</td>
<td>(0.60)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Q23.</td>
<td>2.4</td>
<td>2.5</td>
<td>2.4</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td>(0.98)</td>
<td>(0.94)</td>
<td>(1.12)</td>
<td>(1.15)</td>
</tr>
</tbody>
</table>

**": \( p < .01 \), *": \( p < .05 \), +: \( p < .10 \)

For question 4 (The instructor’s speaking style was easy to listen.) there was a significant interaction between the F-ILS factor and the speed factor, \(F(2, 146) = 3.55, p < .05\). An analysis of simple main effects indicated significant differences for the F-ILS factor and the speed factor. For the F-ILS factor, the Bonferroni test indicated that verbal learners gave significantly higher scores to 1.5× speed than original speed, \( p < .10 \), and original speed than 2.0× speed, \( p < .05 \). On the other hand, visual learners gave significantly higher scores to original speed and 1.5× speed than 2.0× speed, \( p < .05 \). For
the speed factor, the Bonferroni test indicated that verbal learners’ scores for 1.5× speed and 2.0× speed were significantly higher than visual learners’, \( p < .05 \).

For question 11 (I found it difficult to understand the instructor’s voice.), there was a significant interaction between the F-ILS factor and the speed factor, \( F(2, 146) = 2.40, p < .10 \). An analysis of simple main effects indicated significant differences for the F-ILS factor and the speed factor. For the F-ILS factor, the Bonferroni test indicated that verbal learners gave significantly higher scores to 2.0× speed than original speed and 1.5× speed, \( p < .05 \). On the other hand, visual learners gave significantly higher scores to 2.0× speed than 1.5× speed, \( p < .05 \), and 1.5× speed than original speed, \( p < .05 \). For the speed factor, the Bonferroni test indicated that visual learners’ scores for 2.0× speed were significantly higher than visual learners’, \( p < .05 \).

For question 15 (The images displayed were pleasant to view.), there was a significant interaction between the F-ILS factor, \( F(2, 146) = 5.26, p < .05 \). An analysis of simple main effects indicated significant differences for the F-ILS factor and the speed factor. For the F-ILS factor, the Bonferroni test indicated that verbal learners gave significantly higher scores to original speed and 1.5× speed than 2.0× speed, \( p < .05 \). On the other hand, visual learners gave significantly higher scores to original speed than 1.5× speed, \( p < .10 \).

For question 23 (The slides contained many figures and tables.), there was a significant interaction between the F-ILS factor and the speed factor, \( F(2, 146) = 3.42, p < .05 \). An analysis of simple main effects indicated significant differences for the F-ILS factor and the speed factor. For the F-ILS factor, the Bonferroni test indicated that visual learners gave significantly higher scores to 2.0× speed than original speed, \( p < .10 \). For the speed factor, the Bonferroni test indicated that visual learners’ scores for 2.0× speed were significantly higher than visual learners’, \( p < .05 \).

4. Discussion & Conclusion

The goal of this study was to clarify effects of students’ learning styles on their learning experience with lecture videos played at different playback speeds. We focused on the interaction between students’ learning styles and video playback speed. In our experiment, participants’ learning styles were categorized by F-ILS and 35 verbal students and 40 visual students learned about the network infrastructure with lecture videos played at original speed, 1.5× speed, and 2.0× speed.

The comprehension test results indicated that the video playback speeds and the students’ learning styles did not influence the comprehension test scores. The subject evaluation results indicated that there were significant interactions for their learning experience. Especially, visual learners felt much cognitive load when they watched video at 2.0× speed while verbal learners not.

These findings suggested that the possibility that students’ learning experience could differ from their learning styles when they learned something with hi-speed lecture videos. The present data offered further corroboration of the findings of Nagahama and Morita (2017). However, the results are therefore limited to Japanese students, and they need to be replicated with other populations.

References

Word Error Rate as a Listenability Index for Learners of English as a Foreign Language

Katsunori KOTANI* & Takehiko YOSHIMIb

aSchool of International Professional Development, Kansai Gaidai University, Japan
bFaculty of Science and Technology, Ryukoku University, Japan

*kkotani@kansaigaidai.ac.jp

Abstract: In learning/teaching English as a foreign language, it is necessary to prepare listening materials that match learners’ proficiency. Recent development of computer-assisted language learning/teaching solves a matching of proficiency by automatically measuring the ease of listening comprehension (listenability). Previously, an index for listenability was determined by learners’ subjective judgment for listening comprehension. The present study proposed word error rate (WER) in transcription as an alternative listenability index. The experimental results demonstrated the reliability and validity of the WER as a listenability index.

Keywords: listenability, word error rate, learning material, English as a foreign language

1. Introduction

An advantage of computer-assisted language learning/teaching is the use of listening materials taken from the Internet, which results in a heavy burden on language teachers to check whether materials are appropriate for the proficiency of their learners so as to prevent decreases in learning motivation. A solution is to use an automatic measuring method for the ease of listening comprehension (listenability) (Kotani & Yoshimi 2016, Yoon et al. 2016). These previous studies determined an index for listenability by learners’ subjective judgment for listening comprehension on a five-point Likert scale. Although this approach succeeded in measuring the listenability at the text level (Yoon et al. 2016) and at the sentence level (Kotani et al. 2016), it remained open for the possibility of measurement in more detail by measuring the listenability at the word level within context.

Therefore, the aim of the present study is to propose word error rate (WER) in transcription as another listenability index. This study also reports experimental results for the reliability and validity of WER as a listenability index, which was examined within the classical test theory (Brown 1996). The results showed that the WER was as reliable a listenability index as subjective judgment, and that the WER was a more valid listenability index than subjective judgment.

2. Compilation of Listenability Data

Listening materials were produced based on two texts distributed by the International Phonetic Association (1999), and the texts included all of the English phonemes. The voice actor (female, 35 years old, Canadian) read the texts aloud with an American accent at natural speech rate (approximately 180 words per minute (Rodero 2012)).

The listenability data were compiled from 50 learners of English as a foreign language at university (28 males, 22 females; mean age: 20.8 ± 1.3 years old who were compensated for their participation. All learners were asked to submit valid scores (10–990) from the Test of English for International Communication (TOEIC) in the current or previous year. In our sample, the mean TOEIC score was 607.7 ± 186.2.

WER data were derived based on evaluation results of learners’ transcription by two university English teachers: (i) correct transcription, (ii) deletion, (iii) substitution, and (iv) addition, which were annotated using the UAM Corpus Tool (O’Donnell 2008). WER was calculated by dividing the number of transcription error tags (deletion, substitution, addition) by the total number of words in a reference
spontaneous spoken sentence. In order to examine the inter-evaluator reliability, correlation analysis was performed between an evaluator’s WER (WER-A, where the mean value was $0.59 \pm 0.05$) and the other’s WER (WER-B, where the mean value was $0.56 \pm 0.05$), which showed strong correlation ($r = 0.97$). Subjective judgment data were derived from scores subjectively determined by learners on a five-point Likert scale (from 1: easy, 2: somewhat easy, 3: average, 4: somewhat difficult, or 5: difficult), following previous research (Kotani et al. 2016, Yoon et al. 2016).

3. Assessment of the WER as a Listenability Index

The reliability of the WER was examined using Cronbach’s alpha (Cronbach 1970) defined in equation

$$\alpha = \frac{k}{k-1} (1 - \frac{1}{k} \sum_{i=1}^{k} S_{ii}^2)$$

where $\alpha$ is the reliability coefficient, $k$ is the number of items (here: sentences), $S_{ii}^2$ is the variance associated with item $i$, and $S_T^2$ is the variance associated with the sum of all $k$-item values). Cronbach’s alpha reliability coefficient ranges from 0 (absence of reliability) to 1 (absolute reliability), and empirical satisfaction is achieved with values above 0.8. The reliability coefficients for WER-A, WER-B, and the subjective judgments (WER-A: $\alpha = 0.97$, WER-B: $\alpha = 0.97$, and subjective judgment: $\alpha = 0.89$) outperformed the baseline.

The validity of WER was examined in terms of whether it reflected learner proficiency-dependent listenability. The dependence of listenability on a learner’s proficiency refers to the situation in which listenability is higher for learners at a high proficiency level than for those at a low proficiency level. Along with TOEIC scores, mean WER and subjective judgment values of learners were calculated by dividing the sum of WER/subjective judgment values by the number of sentences (15 sentences).

The construct validity of WER was examined from the viewpoint of the distinctiveness between proficiency levels: beginner (TOEIC scores ranging from 295–450), intermediate (490–685), and advanced (730–900). Table 1 shows the mean (SD) values for WER-A, WER-B, and subjective judgment of these groups.

<table>
<thead>
<tr>
<th></th>
<th>Beg. (n = 16)</th>
<th>Int. (n = 16)</th>
<th>Adv. (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WER-A</td>
<td>0.69 (0.17)</td>
<td>0.64 (0.17)</td>
<td>0.40 (0.19)</td>
</tr>
<tr>
<td>WER-B</td>
<td>0.68 (0.17)</td>
<td>0.62 (0.17)</td>
<td>0.38 (0.19)</td>
</tr>
<tr>
<td>Subjective judgment</td>
<td>4.5 (0.7)</td>
<td>4.3 (0.7)</td>
<td>3.9 (0.9)</td>
</tr>
</tbody>
</table>

One-way ANOVA showed statistically significant differences ($p < 0.01$) between all three groups of learners for TOEIC scores ($F (2, 47) = 235.4$), WER-A ($F (2, 47) = 39.7$), WER-B ($F (2, 47) = 41.6$), and subjective judgments ($F (2, 47) = 9.2$). The results were further examined using Tukey’s post-hoc comparison, which showed statistically significant ($p < 0.01$) in the WER and the subjective judgment between the beginner and advanced levels and between the intermediate and advanced levels, but not between the beginner and the intermediate levels.

The criterion-related validity of WER was examined from the viewpoint of the correlation with learners’ TOEIC scores. TOEIC scores showed strong correlations with WER-A ($r = -0.83$) and WER-B ($r = -0.84$), but a weak correlation with subjective judgments ($r = -0.51$). According to the TOEIC technical manual (Chaucney Group International 1998), empirically valid correlation coefficients above 0.73 were found, as TOEIC scores were correlated with a valid English test ($r = 0.73$). The WER and subjective judgment were further examined in an asymptotic $z$-test with by using Fisher’s $z$-transformation (Lee & Preacher 2013). Statistically significant differences ($p < 0.01$) were observed with WER-A ($n = 50, z = -3.0$), and with WER-B ($n = 50, z = -3.2$). That is, the WER was more valid than subjective judgment.

4. Conclusion

This study examined listenability indices by comparing subjective judgment of previous studies (Kotani et al. 2016, Yoon et al. 2016) with WER of transcription proposed by this study. The reliability and...
validity of the WER, assessed using classical test theory, were confirmed as well as the subjective judgment. WER outperformed the subjective judgment in the criterion-related validity. This high criterion-related validity seems to be caused by the objective evaluation which excludes over/underestimation by learners.

The experimental results support the use of a listenability measurement method in a classroom, where a listenability measurement method is available as an education application. First, a teacher/learner picks up listening materials on the Internet. Second, listening materials are examined linguistically in order to extract linguistic features such as sentence length and speech rate. Third, linguistic features are input to a listenability measurement application, in addition to a learner feature of listening proficiency in terms of a TOEIC score. Then, the application calculates and demonstrates the listenability of the materials. According to the results, a teacher/learner chooses listening materials among three types of listenability: low, moderate, or high. Materials with high listenability should be chosen for extensive listening practice, and those with low listenability for intensive listening practice.

A remaining problem of this study is to seek another alternative listenability index that consists of both subjective and objective evaluations. Subjective judgment is plausible in that it directly demonstrates learners’ listenability, which is difficult for WER to explain. The problem of subjective evaluation would be solved by combining with objective evaluation, that is, WER. Future study needs to assess the reliability and validity of a complex listenability.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. All remaining inadequacies are ours alone. This work was supported by JSPS KAKENHI Grant Numbers, 22300299, 15H02940.

References

Science Teachers’ Engagement with ICT in Singapore: Different Perspectives

Aik-Ling TANa & Seng Chee TANb

aNational Institute of Education, Nanyang Technological University, Singapore
bCRADLE, Nanyang Technological University, Singapore

*aikling.tan@nie.edu.sg

Abstract: In this paper, we present narratives of three in-service biology teachers in their journey with the use of Information and Communication Technology (ICT) in their practices. These narratives provide useful insights into the in-service science teachers’ ideas, dilemmas and actual usage of ICT. The use of narratives to present perspectives of biology teachers’ engagement with ICT is a deliberate one — different teachers have different experiences with their students and across different schools. As such, the stories generated are different but personal and real to each participant. The in-service teachers are purposively sampled: all of them having taught science in secondary school for at least three years and had shown a keen interest in technology during their pre-service teacher program. A series of questions was used to help the participants reflect on their experiences and craft their narratives. These narratives were then analysed using content analysis of recurring themes. From the narratives, we found that generic ICT tools could be used for evaluation of students’ learning while specific ICT tools such as sensors were used for the teaching of specific scientific concepts and to support scientific inquiry. Further, in deciding which ICT tool to adopt for their lessons, teachers took into consideration external factors such as availability of wireless networks, school infrastructure, ease of setting, and students’ motivation. In terms of professional development on the use of ICT, we found that sharing sessions on what works, time and space for experimenting with new ideas, and in-depth implementation of fewer ideas rather than many ideas worked for the teachers.

Keywords: In-service science teachers, Narratives, ICT use, Professional development

1. Introduction

The use of Information and Communication Technology (ICT) in education has been ubiquitous for more than two decades. Information and Communication Technology has changed classrooms and the way people learn faster than any other educational reforms or initiatives. For example, in science teaching and learning, the call for teaching science as a form of inquiry started in the 1900s, and until today, the adoption of science as a form of inquiry remains fuzzy and sometimes even elusive. The penetration of ICT into science classrooms, however, appears to be more successful as evident from the pervasive use of laptops, sensors, videos, mobile computing devices and apps for science learning. The popularity of ICT in science learning points to the need for science teachers to be proficient and confident in their use of ICT to enhance the teaching of science as a form of inquiry. As Singer and Maher (2007) pointed out, despite the prominence of learning technology in schools, significant changes in teaching strategies have not kept pace. They attributed this to the poor quality of professional development and inadequate teacher preparation. They argued that since the presence of ICT tools is characteristic of the 21st-century classroom, teachers need to know how to use computers, software and apps. In science classrooms, ICT can help students conceptualize invisible and abstract science processes, visualize particles that are either too small for the naked eyes or too large for normal sight, and to assemble scientific model and propose explanations (George & Ogunniyi, 2016).

A positive attitude of teachers towards ICT integration and their willingness to try various tools are important for effective use of these tools in the classroom. (Cracium & Bunoiu, 2015). Cracium and Bunoiu further argued for the need to relate the digital skills possessed by in-service and pre-service teachers be made sense of in the light of “their attitudes towards ICT in their everyday life for personal
and professional usage…” (p.82). The call for more attention to the needs and attitudes of teachers is heeded in this paper as we focus on science teachers’ personal ideas and experiences with ICT guided by the research question: “What are the ICT experiences of science teachers in teaching and learning of science?

2. Background Literature

In this review, we raised some questions at the nexus of science teacher professional development and the use of ICT. From the literature that we have read, we attempt to show that there is a need for us to better examine teacher professional development in ICT from a more teacher-centric angle. Science education journals published few articles pertaining to the use of ICT and teacher professional development. The articles that are published with respect to ICT are generally focused on how students respond to the use of ICT to learn specific science concepts. The only exception is the Journal of Science Teacher Education that revealed more articles related to how ICT is used to improve professional development and pre-service teacher education. Technology-related journals published articles related to science and use of ICT as well with a distinct difference from the articles published in science education journals. The articles published in technology-related journal tend to foreground the technology with little mentioned about the specific science content. After a scan of the literature from science education journals and technology-related journals, two key questions were raised.

The first question that arouses our curiosity pertains to what research has already informed us about effective professional development for science teachers in the integration of ICT to enhance science learning. Research findings of various studies generally point to a co-design modality as an effective way of equipping teachers with technology skills (Kafyulilo, Fisser, & Voogt, 2015; Leary, Severance, Penuel, Quigley, Sumner, & Devaul, 2016; Matuk, Gerard, Lim-Breitbart, & Linn 2016). Teachers working together with technological experts and researchers to co-design lessons not only enables teachers to increase their understanding of the relationship between the science content, the technology, and the lesson design but also empowers them to make critical decisions about what is best for their students’ learning. This is powerful in encouraging meaningful and sustainable change in classrooms.

The second question that we asked relates to the methods that are currently used to understand teachers’ needs and ideas about ICT and science learning. George and Ogunniyi (2016) used a survey (Technology Acceptance Model) and interview with 45 teachers in South Africa to determine the usefulness and intention of teachers to use ICT in their practices. Craciun and Bunoii (2015) used a self-report evaluation on a group of pre-service teachers to assess change in confidence level after being engaged in using social media and Web 2.0 technology for digital story-telling and digital comics in science. Qablan, Abuloum, and Al-Ruz (2009) used a series of interviews and classroom observations together with Activity Theory to assess the utilization of ICT in science teaching in Jordanian schools. Chittleborough (2014) studied chemistry pre-service teachers’ produced artefacts, their reflections and survey to determine if integrating technology into pre-service teachers programs will help to increase skill levels of use of technology for teaching and learning. The literature revealed few studies that used a research method that enables one to delve deeply into the experiences of teachers as they engage with ICT. The lack of teachers’ voices in the use of ICT is one area that we aim to address in this paper.

3. Methods

This is a qualitative study that adopts a narrative approach to studying the lives of science teachers as they engaged in their business of teaching and learning. Purposive sampling of the three participants was carried out. The participants are all male in-service teachers who have been teaching science and biology for at least three years in a secondary school. These teachers are chosen as they have been shown to be confident users of technology while they were pre-service teachers. The teachers were invited to write a personal narrative of their experiences with ICT in their teaching. Prompts in the form of questions (such as what factors do you take into account when you plan to use ICT) were given to guide their reflection. We read through the narrative and surfaced questions that arose and send these to
the participant teachers to answer. We argue that working with the participant teachers on their experiences with ICT usage enables both the participants and us (as researchers) to become more sensitive and curious about our experiences. The research process in itself is a powerful learning experience that cannot be derived from books or formal training (Teo & Tan, 2011). Through this process, we refine their thoughts and our interpretation of the ideas.

The use of narratives as a means for us to present the different experiences that science teachers have with using ICT in their practices is a deliberate one. Narratives are stories that people tell based on their personal experiences with the world around them. From these stories told, it revealed how individuals make sense of and interpret the world around them. As such, using stories, we can understand the existence of different interpretations of social life (Clandinin & Connelly, 2000) – in this case, how the policies of ICT and its local implementation are experienced by teachers. Personal narratives are able to reveal more personal perspectives of the world than interviews or questionnaires can. As Nicole Grimes (2013) wrote in her personal narrative of her identity as a Caribbean female teaching science, she justified her choice of using narratives by arguing that “the stories we craft describe our perceptions and our experiences, and in themselves are highly significant, as when examined closely, they provide us with information about our human culture” (p. 334). In this study, we present narratives from three secondary science teachers in an attempt to understand the role that ICT play in the teaching and learning of science.

Coding was also carried out on the three narratives. Although the sample of three is small, the coding was meant to under the frequencies with which external factors and internal factors affecting their experiences with the use of ICT in their practices were mentioned. Coding was carried out in NVivo and the coding categories were emergent from the narratives.

4. Results

In this section, we present the three narratives written by three science teachers, Glen, Edmund and Hugo (pseudonyms). The narratives themselves are stories that ought to be read, reflected and interrupted by the readers, sometimes in the context of the readers’ personal experiences. As such, the narratives, in themselves, form a rich corpus of information from which we can learn from. In the second part of this section, we presented the frequencies with which the various external and internal factors were mentioned in the three narratives to highlight the emphasis that the teachers places of each of the factors.

4.1 Glen: “How technology can allow AfL to happen in the classroom.”

I am a biology teacher who has been teaching for 10 years. I had previously taught at junior college students and now is teaching secondary school students. From the many years of teaching biology, integrating ICT into my practice means that I use ICT (e.g. PowerPoint software, Edmodo, Socrative and Nearpod apps) to support my teaching and students' learning. I have recently changed my style of teaching in the science classroom. In the past, I used to teach using PowerPoint slides and ask questions verbally in class to check on students' understanding. When I finish teaching the entire topic, I will give students worksheets to complete and then I will go through the answer with them in class. Nowadays, I have used Nearpod for my lessons. I upload my slides into the app and integrate the questions from my worksheet into the app. Once I finished a concept in a topic, which can be about 2-3 slides, I will show questions related to the concept and get students to respond immediately on their mobile devices. I will be able to use the data collected to give on-the-spot feedback to close any learning gaps the students may have, before proceeding to teach the next concept. ICT has also allowed me to do flipped learning as I would upload some video links, etc. on Edmodo and students will learn about those concepts first before attending lessons. So the shift in my use of ICT in these 10 years is from simple PowerPoint slides (presentation, frontal) to Socrative (elicit responses) to Nearpod (presentation with eliciting responses).

The availability of ICT tools has influenced how I planned my lesson. Nowadays, I need to prepare all the questions before I teach a topic so that I can integrate the questions with my slides as I go through Nearpod. In the past, I would prepare and give the questions after I have finished teaching a
topic. I need to be mindful about which ICT tool I use in my lessons, especially if I collect responses from students for the purpose of formative assessment. Every ICT tool has its strength and limitations, and a good awareness of this is important in my selection. For example, Kahoot is easy to set up but only allows multiple choice items; Nearpod allows students to provide open-ended responses, but it takes more time to prepare.

With respect to the role that ICT play in science teaching, I don't really distinguish between using ICT in science teaching or the teaching of other subjects. The choice of ICT tool should be fit-for-purpose. The only benefit of using ICT for the teaching of biology is that simulations can be shown to help students better visualise processes, e.g. transcription and translation, at the cellular/molecular level.

Beyond science learning, I think ICT enhances students' learning in general because I use it a lot for assessment for learning (AFL). Technology has now allowed me to collect data on students' learning on-the-spot and I can adjust my pace of teaching and tailor my delivery to better suit the profile of learners. I use the information collected with the help of ICT tools to inform my next pedagogical move in the classroom.

The decision to select the most suitable ICT tools for my lessons rely on different factors such as the (1) infrastructure (whether the school's internet wifi is able to support), (2) time to set up the equipment (I settled with Nearpod because it is relatively quick to set up), and (3) student motivation (I have not used it for my NT class because they tend to be distracted by mobile devices). I am trying to explore other ICT tools (e.g., Plickers) for this group of students such that I can gather quick responses without having them to have any computers/mobile devices.

In my school, the school leaders have dedicated staff training days for staff to learn ICT tools. For example, in the training session last year, I learnt about Edmodo. This year, the EdTech team taught us Nearpod, Kahoot, and Padlet. I have chosen to explore further into Nearpod and am intending to embark on my own research about the use of such apps to support teaching and learning. I think more research needs to be shared with teachers about how an ICT lesson could be structured and how technology can allow AFL to happen in the classroom.

4.2 Edmund: “Grounds up construction of knowledge with much convenience.”

I have been teaching in a school for 3.5 years. For me, the integration of ICT ranges from the basic use of a projector to the more advance use of motion analysis app such as Tracker (http://physlets.org/tracker/). For example, in the teaching of how objects fall (see attached for lesson plan), a video can be made of a tennis ball that was dropped from a certain height. This can then be analysed using Tracker to plot various graphs such as the velocity-time graph, and calculations can be done to determine the acceleration as the tennis ball falls. Results from such analysis can then be used to reinforce concepts taught in Kinematics.

When I plan for my science lessons, my primary concern still lies in how the content can be delivered to the students in a manner that ensures the accurate learning of that particular concept. I will only use ICT if it aids in the teaching and learning of the subject matter. That being said, the basic use of ICT tools such as PowerPoint slides, animations, and videos are now indispensable in my day to day lessons as they allow for more clarity of thought in the ideas that are being communicated to the students. Generally, the factors that I will consider include whether the ICT tool is supported by the school's infrastructure, the cost of the ICT tool, the 'user friendliness' of the ICT tool (since a class of 40 students with diverse backgrounds may need to handle the ICT tool), how does the ICT tool brings about learning, the time spent teaching on the use the ICT tool vs. the actual time spent learning the concepts by students.

Beyond running simulations for a richer and easier way of teaching scientific concepts, ICT can bring about the much-needed authenticity into the teaching of science. The affordance of cheap sensors such as those in Arduino allows for experimental data to be gathered readily from the environment and analysed to verify scientific laws. This removes the abstractness of ideas in the teaching of science and helps students to reconcile their day to day experiences with the concepts that are being taught.

From my experiences of experimenting with the various ICT tools such as Tracker, Arduino, and others, the use of ICT tools definitely helps to enhance science learning as it allows for a grounds-up construction of knowledge with much convenience. Previously, such construction of knowledge requires students to carry out of experiments in science labs, which are not only time and
resource intensive, they are also more prone to mistakes by students. However, with videos, animations, and simulations, the scientific inquiry process can be made more efficient and meaningful with more focus on the discussions of ideas by the students. Sensors can also be readily deployed to not only gather large amounts of experimental data to minimise any random errors, but also to directly verify scientific laws (for example, acceleration is directly measured using accelerometers rather than derived from the double differentiation of a displacement-time graph which is gathered from the use of a ruler and a stopwatch).

When I first started out, there was a lot of usage of videos and simulations (https://phet.colorado.edu/), but now, what interests me are the apps and sensors that can allow me to directly verify scientific concepts to bring authenticity into the classroom. To help me use ICT tools more effectively, I think the sharing of lesson ideas for the integration of ICT would be great since most teachers are more interested in how ICT tools can help elucidate scientific concepts rather than the ICT tools themselves.

4.3 Hugo: “We have tried too many things.”

I am a teacher who has been teaching Biology in a secondary school for the past three years. My students are keen to learn, and they are generally well behaved. My school is supportive of innovative practices in teaching and learning. For me, integration of information and communication technology (ICT) suggests a pervasive use of ICT in teaching and learning that can be sustained regardless of context. It could be as simple as using a routine method of assessment for learning. For example, I frequently use the powerpoint app "polleverywhere" as a method of understanding my students’ knowledge before the start of lessons. Students use their mobile phones to answer questions in real-time. Other than having a real-time understanding of the data, I can collate the data back-end to compare if they have improved over the course of the lessons.

The use of ICT does influence how I plan my lessons. For me, the use of ICT makes teaching and learning more efficient. I constantly think about how best to make my understanding of my students’ learning efficient. This includes the use of ICT to gather data to capture students learning.

Personally, I do not think there is a specific role of ICT in science teaching, but in teaching as a practice. There might be ideas that seem more "applicable" to the context of teaching science, such as the use of visualisations or animations (on the PhET websites). Beyond that, I would like to zoom out of the science context and look at the bigger picture and ask myself "why ICT in my teaching?" In science, ICT helps to make the learning of abstract concepts more efficient. Through visualisations and animations, it is possible to visualise concepts that are otherwise not feasible to demonstrate or to manifest in a classroom context. However, it is only with clever and thoughtful planning supplemented with instructions/worksheets before learning can be done. So the question is — are we able to have a really immersive ICT lesson with interaction with only an ICT tool with real-time data collection, without the need for pen and paper? Would that then allow me to enhance the learning of science?

In considering if I should use ICT, I think efficiency and scalability are two very important factors. There is little point in using ICT if it means more work with little impact. In my opinion, ICT should make the job easier, and not tougher. I am beginning to use more ICT in my science teaching but have a preference for ICT that builds into routines that engage students. Importantly, I would also use ICT that allows me to have a measurable impact in order to understand how students have learnt or are weak at.

With regard to professional development needs for myself related to ICT, I think we have tried way too many things. I think we should be using a few things more often in class than to keep trying new things. But if there is one that I would like to try it will be a long-term collaborative tool. Till now, I do not think any Learning management system is able to engage students for a long time in order to learn.

4.4 Factors Shaping Science Teachers’ Experiences with ICT

The factors that shape the participant teachers’ experiences with ICT can largely be understood as internal factors (teachers’ beliefs, teachers’ competencies) and external factors (cost, ease of usage, infrastructure, hardware, software, perception of students, availability of time). Table 1 reports the
frequencies in percentage of the mention of the various factors in the three narratives. We hypothesize that the prevalence of mention of any factors is indicative of the concerns or importance that the teachers placed on the factors. In their narratives, the teachers make reference to what they believed to be important functions and purposes of incorporating ICT in their practices. All the three held the belief that ICT ought to make learning more effective and authentic. They however, deferred in the kinds of ICT tools that they would use to achieve their purposes. The high frequency of mention of different types of softwares and Apps is indicative of their different experiences and forms of interaction with ICT tools.

Table 1: Frequencies of factors shaping teachers experiences with ICT

<table>
<thead>
<tr>
<th></th>
<th>Percentage (Frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>External Factors</strong></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>6</td>
</tr>
<tr>
<td>Cost</td>
<td>6</td>
</tr>
<tr>
<td>Perception of students</td>
<td>2</td>
</tr>
<tr>
<td>Availability of hardware</td>
<td>6</td>
</tr>
<tr>
<td>Ease of usage</td>
<td>8</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>4</td>
</tr>
<tr>
<td>Exposure to software</td>
<td>20</td>
</tr>
<tr>
<td><strong>Internal Factors</strong></td>
<td></td>
</tr>
<tr>
<td>Teachers’ beliefs</td>
<td>38</td>
</tr>
<tr>
<td>Teachers’ competencies</td>
<td>0</td>
</tr>
<tr>
<td>Strategies for teaching</td>
<td>10</td>
</tr>
</tbody>
</table>

*Total number of codes is 50

5. Discussion

The narratives of the three participant teachers on their ICT experiences in science classrooms revealed that the same ICT policy has yielded similar yet different experiences for the teachers. Each of the teachers has slightly different reasons for using ICT tools in their teaching. For Glen, his key reason for using ICT tools is to obtain feedback on his students’ progress of learning. This is one of the technological pedagogical knowledge (TPK) – use digital technologies to help in assessing students learning (Graham, Burgoyne, Cantrell, Smith, St. Clair, & Harris, 2009). For this purpose, he does not focus entirely on tools pertaining to science learning. Rather, the ICT tools that he used are generic tools that enable him to ask questions and obtain answers from the students. Hugo shared similar application of ICT tools in his practice. He also used ICT tools to capture students’ learning. The use of ICT tools for evaluation of students’ learning is one that is important for both Glen and Hugo, and this resulted in Hugo wondering if there is a need to distinguish between usage of ICT in science teaching as compared with teaching in general. He is of the opinion that there is no one “specific role of ICT in science
teaching, but in teaching as a practice.” The use of ICT within specific domains was also an area of interest among researchers that lead to the development of the constructs of Technological Pedagogical Content Knowledge (TPACK or TPCK) (Graham, et. al, 2009).

Edmund, reflecting on his experiences with ICT in science teaching, focuses on tools that are directly related to science, that is, his focus is on using technological content knowledge (TCK). TCK emphasizes the use of digital technologies in the data collection, recording and analysis in science. He describes in great detail about using sensors to collect data as part of the science inquiry process and also about the use of tracker app to enable students to visualize motion. The emphasis on these ICT tools to collect data is a result of his goal in his teaching – to ensure that his students learn the content of science. It is evident that Edmund’s focus of ICT is the process of teaching and learning rather than evaluation. Edmund used the affordances of various ICT tools to ensure that the learning of science concepts is made more efficient and authentic. The authenticity comes from the opportunities presented to students to work with data and process them to be used as evidence in their explanations.

We have seen how the three teachers can potentially use ICT for either evaluation or teaching and learning, that is, teachers apply the whole range of knowledge from Technological Pedagogical Content Knowledge (TPACK) to Technological Pedagogical Knowledge (TPK) to Technological Content Knowledge (TCK). Beyond these curriculum matters, when teachers are deciding on whether to use ICT in their science lessons, they also take into account external factors such as the availability of wireless connections, the school infrastructure, the amount of time required for setting up and the students’ abilities. Glen and Edmund expressed the need to consider these external factors explicitly while Hugo discussed these factors in the light of increasing efficiency of teaching. These concerns about the factors that will promote the use of ICT is similar to what George and Oyunniyi (2016) found from their study using the Technology Acceptance Model. They found that factors such as the ease of usage and the usefulness of the ICT tools are important considerations for teachers. Similar findings with regard to frustrations of insufficient and unreliable ICT infrastructure, when needed, were also reported by Qablan, Abuloum, and Al-Ruz (2009).

From the narratives, it is clear that the three participant teachers have different views with regard to professional development to improve their use of ICT tools in science teaching and learning. For example, Hugo believes that he has been introduced to too many ideas to try in his classroom. He would prefer to test fewer new ideas and have more in-depth experimentation with the ideas. Edmund thinks sharing of ideas of what works in ICT is a useful and practical way to learn new tools. For Glen, he prefers to learn about new ICT tools from formal structured sharing and subsequently be given to exploring its implementation in his practices. He also thinks that more research based evidence of ICT tools ought to be shared with teachers.

6. Conclusions: Implications for Science Teacher Professional Development

In this paper, we set out to examine the experiences of science teachers as they engage with ICT tools in their practices. The three narratives presented shows how three different teachers perceived their ICT experiences in teaching science with the same national ICT policy but interpreted by their personal preferences and opinions. Two key areas emerged from the narratives on how ICT tools are used in science teaching — used of generic ICT tools to obtain feedback on students’ learning and use of more science-specific tools to collect data and engage students in the authentic inquiry of science. As these are areas that are probably of interest to science teachers, professional development on the use of ICT tools for science teachers can draw explicit examples to how the various tools such as used of sensors, digital storytelling, videos, trackers and others can be used to support scientific inquiry and also to create avenues for collecting feedback on students learning. It is of value to draw reference of how the various tools can be used to support the learning of specific scientific concepts. Hence, we argue for the need for explicit instruction of concepts and principles of TPACK, TCK and TPK in teacher professional development and pre-service teacher education. Matuk, Gerard, Lim-Breitbart, and Linn (2016) presented evidence to support this idea that technology can help to enhance teachers’ understanding of patterns of students’ thinking, manage class progress at individual or group levels and acquire information to aid teachers in making modifications for their next pedagogical move (AfL).

Secondly, the narratives revealed that science teachers also take into account the ease of usage and implementation in considering the choice of ICT tools. We argue that this implies that the
successful use of ICT is not merely the business of the teachers or isolated in the classroom alone. Rather, there has to be an ecosystem to support the teachers and their implementation of ICT to enhance their teaching and learning. Administrators can ensure that school infrastructure is well developed to support ICT use. Staff involved in technical support of ICT should also be well trained to help with trouble shooting and to solve technical issues related to implementation. This will ensure that teachers have time to devote to the teaching and learning aspect of technology. The intention to use technology can be influenced by various variable such as perceived ease of use, perceived usefulness, attitude toward technology and motivation (George & Ogguniyi, 2016).

Lastly, the modalities of professional development for science teachers related to ICT should be varied to meet the different demands and needs of teachers. From just three narratives, we saw three different opinions of what professional development ought to be (ranging from sharing of ideas on what works to giving time for exploration after workshops, to creating space to try new and fewer ideas instead of many ideas). The narratives indicated that professional development also need to be varied to tackle different forms of knowledge related technology (TPACK, TPK, and TCK) use in schools. As such, professional development providers need to offer a more sincere listening ear to understand different groups of teachers and the stages of their journey with ICT before offering ICT solutions for their classrooms. Different modalities of professional development can also be adopted. For example, Matuk, Gerard, Lim-Breitbart, and Linn (2016) showed that by getting teachers to work on various technological tools together with researchers helps to empower them to design better and more meaningful lesson. Moreover, the teacher will be able to develop a better understanding of the relationship between technology, teaching, and design. Similarly, Leary, Severance, Penue, Quigley, Sumner, and Devaul (2016) also adopted a co-design mode to help teachers appreciate how to better resolve the tensions between the Next Generation Science Standards (NGSS) and their existing classroom practices. Kafyulilo, Fisser, and Voogt (2015) also provided positive evidence in support of a co-design mode of teacher professional development for the use of ICT.

Professional developers could also possibility move beyond the technology and foreground the scientific content to help science teachers better appreciate the usefulness of the technology in fulfilling science as a form of inquiry. Earle (2002, p. 8) aptly argues that “integrating technology is not about technology — it is primarily about content and effective instructional practices. Technology involves the tools with which we deliver content and implement practices in better ways. Its focus must be on curriculum and learning. Integrations is defined not by the amount or type of technology used, but by how and why it is used.” We support this argument and hence perhaps the construct of Technological Pedagogical Content Knowledge (TPACK) could be better utilized to fulfil this. On the other hand, we could also question if there is a venue for the professional development of science teachers into the epistemology of technology, that is, understanding ICT as the content itself and as a way of knowing (van Eijick & Claxton, 2009).

Acknowledgements

We would like to thank the three participate teachers who have patiently worked through their personal narratives with us so that we are able to understand their experiences better.

References


Participation and Psychological Ownership on Teachers’ Beliefs of a Cloud-based VLE

Joanne Sau-Ching YIM ab*, Priscilla MOSES a* & Alia AZALEA a
a Universiti Tunku Abdul Rahman, Malaysia
b Tunku Abdul Rahman University College, Malaysia
* yimsc@tarc.edu.my, * priscilla@utar.edu.my

Abstract: Perceived usefulness and perceived ease of use constitute important belief factors when technology adoption decisions are made within a non-mandatory setting. This paper investigated the roles played by users’ participation and psychological ownership on these beliefs about using a virtual learning environment (VLE). Participation is conceptualised as the extent of activities and behaviours teachers exhibit in a VLE, while psychological ownership refers to the feelings of ownership developed for it. As technology become ubiquitous, psychological ownership is increasingly becoming a relevant phenomenon in technology adoption research, where people can feel psychologically attached to a particular technology. It is proposed that such phenomenon can also occur when using a VLE. Hence, hypothesised relationships were tested with 67 Malaysian teachers from three schools who are using a cloud-based learning platform. Measures were adopted and adapted from established scales used in previous studies. Results from partial least square analysis found significant effect of participation on psychological ownership ($\beta = 0.661, p < 0.001$), while psychological ownership had significant effect on perceived usefulness ($\beta = 0.589, p < 0.001$) and perceived ease of use ($\beta = 0.632, p < 0.001$). This study provides initial evidence for participation and psychological ownership to be included in examining users’ belief about educational technologies.

Keywords: Participation, Psychological ownership of VLE, Perceived usefulness, Perceived ease of use

1. Introduction

Cloud-based learning platforms are gaining popularity for its advantages in offering access to infinite on-demand resources, unlimited storage, and scalability in terms of bandwidth and computing functionalities (Stantchev, Colomo-Palacios, Soto-Acosta, & Misra, 2014; Hew & Kadir, 2016). Users are found to prefer this new platform to traditional Learning Management Systems (LMS), rating it favourably in terms of their perceived ease of use, perceived usefulness, and attitude towards the platform (Stantchev et al., 2014). The services offered by cloud-based learning platforms surpass traditional LMS (Shiau & Chau, 2016), with benefits such as the ability to be “omnipresent” to the access of content, and provision of powerful collaborative support (Park & Ryoo, 2013). Despite its advantages, the current cloud based learning platform in Malaysia, the Frog VLE, has received lukewarm application from teachers (Auditor General Report, 2013).

The Frog VLE platform was implemented in 2012 as part of 1BestariNet (1SmartNet) project, a signatory educational initiative to catapult technology-based teaching and learning in Malaysian public schools (Soon, 2014). Through this programme, Malaysia became the first nation in the world to connect all 10,000 public schools, 500,000 teachers, 5.5 million students, and 4.5 million parents using a single, cloud-based learning platform with high speed 4G internet connectivity (Hew & Kadir, 2016). This cloud-based VLE offers virtual equivalents of conventional educational concepts, wherein teaching and learning can be conducted virtually, parents can view their children’s tests results and school news, while the school management can disseminate information via the platform (Soon, 2014). It is integrated with educational applications such as Khan Academy and Google Apps for Education, providing users with an array of functionalities through Widgets and built-in applications. Some of the functionalities include: assignments, e-mail, booking calendar for school resources, creation of learning
sites, quizzes, learning style reports, and Google Drive. There were also links to virtualized community and resources through forums, bookshelf, departmental sites for school subjects, the Pond, FrogStore, FrogAcademy, personal dashboard, and the school dashboard.

The first phase of implementation in 2013 to 2015 focused on infrastructural set-up and the equipment of teachers with VLE competency (Ministry of Education [MOE], 2013). Training of teachers were conducted in a ‘train the trainers’ manner, where teachers who have attended initial trainings would train their colleagues in their respective schools (Cheok & Wong, 2016). The second implementation phase which last from 2016 to 2020 concentrates on reviewing best practices, and using ICT for intervention of specific groups such as rural schools or under-enrolled schools (MOE, 2013). Teachers have cited the major success factors of the VLE depend on its physical attributes, tool for collaboration, and its functions as pedagogical tool (Soon, 2014). Usage of the VLE among teachers remain low, despite the realised benefits offered by the platform in enhancing students’ learning (Cheok & Wong, 2016; Rajaendram, 2017).

2. Literature Review and Objective of the Study

2.1 Perceived Usefulness, Perceived Ease of Use, Participation and Psychological Ownership

The Technology Acceptance Model (TAM) (Davis, 1989) is a theoretically justified and parsimonious model that makes a good ground theory for studying users’ beliefs in using e-learning technology (Šumak, Heričko, & Pušnik, 2011; Teo, Lee, Chai, & Wong, 2009). Perceived usefulness and perceived ease of use constitute the two central variables in the TAM (Davis, 1989). Perceived usefulness (PU) describes the belief that using a particular technology will improve one’s job performance, while perceived ease of use (PEOU) refers to the belief that using a target technology will be free of effort (Davis, 1989). Within the education milieu, these beliefs are found to have pronounced influence on users’ attitude, intention and usage behaviours (Moses, Wong, Bakar, & Mahmoud, 2013; Wong & Teo, 2011). In the present study, PU describes the degree to which teachers believe that using cloud based VLE can improve their teaching performance, while PEOU refers to the extent to which teachers believe such systems can be used easily (Wang & Wang, 2009).

Since its conception, TAM has evolved and extended with different external variables to explain its main constructs (Marangunić & Granić, 2015). When developing one of the final iterations of the model, the TAM 3, it was suggested that user participation be incorporated in future investigations of TAM (Venkatesh & Bala, 2008). User participation originally refers to the extent of activities, assignments, and behaviours that users undertake during the systems development and implementation process (Barki & Hartwick, 1994). However, scholars have advocated the examination of participation beyond the system development phase, to focus on post implementation to understand users’ engagement in the system (Ju, Wei, & Tsai, 2016; Shen, Khalifa, & Almulla, 2013; Wagner & Newell, 2007). The present study was conducted in the end user domain, and end users’ participatory activities are very much different from the developmental phase of a system. The literature has identified two broad categories of activities in system design that involve end-users: parameterization and tailoring (Ardito, Buono, Costabile, Lanzilotti, & Piccinno, 2012). Parameterization describes activities that allow end users to choose among alternative behaviours and mechanism that is available in the system to customize the particular system (Lieberman, Paternò, Klann, & Wulf, 2006). Tailoring involves the creation or modification of a particular software by end users to suit their needs (Ardito et al., 2012). As users become involved in participatory activities in the system, it becomes personally relevant and important to them, therefore inducing their ownership and buy-in of the particular system (Spears & Barki, 2010; Wu & Marakas, 2006). A similar dynamic may also occur in the context of cloud VLE, as one of the main services offered by cloud VLE, the software as a service (SaaS), provides various customizable applications that run through the cloud. These applications and programs such as Google App for Education offers high degree of tailoribility, providing teachers with choices to set parameters to build and design their VLE environment for instructional purposes (Shiau & Chau, 2016). Based on this contention, this study conceptualise participation as the extent of activities and behaviours teachers perform in the cloud VLE, as they tailor the applications and programs which run through the cloud for teaching and learning.

As technology becomes ubiquitous in daily lives, psychological ownership has become a
relevant phenomenon in technology adoption research (Klesel, Ndciu, & Niehaves, 2016). Psychological ownership is the possessive feeling of being psychologically attached to an object, where individuals feel as though an object’s ownership or a part of that object is ‘theirs’ (Pierce, Kostova, & Dirks, 2001). Within the work context, the potential targets of ownership can include tangible and intangible objects, tangible objects can include one’s workspace, or work tools, whereas intangible objects may include the organisation itself or a project that one leads (Pierce & Jussila, 2011). Technology can also be a target object for which psychological ownership can be developed, as evidenced by research carried out in contexts such as in virtual world (Lee & Chen, 2011), clinical information system (Barki, Paré, & Sicotte, 2008), and social media (Karahanna, Xu, & Zhang, 2015).

Psychological ownership has received attention as an important factor in affecting users’ behaviour in technology adoption (Klesel et al., 2016; Lee & Chen, 2011). Pierce et al. (2001) proposed four human needs which give rise to the development of psychological ownership: to have a place, to have effectance, to be stimulated, and to have a self-identity. A cloud-based VLE as a target of ownership can potentially fulfil these needs as it present ample opportunities for teachers to develop psychological ownership for it. As VLE content and spaces are customized by teachers, they may feel psychologically attached to their ideas, design, and intellectual contribution. Such personalization of space promotes a sense of familiarity, satisfying the need of having a ‘place’ (Porteous, 1976). By using and tailoring the applications available, teachers exercise discretion in using resources available to enhance their teaching, satisfying the effectance need. Taking part in the VLE stimulates teachers with different functionalities and applications in the system, satisfying the need for arousal which can explain the dynamics of psychological ownership (Pierce & Jussila, 2011). The flexibility afforded to teachers to define their VLE space can satisfy the need for self-identity, as they express themselves through using the features available. With this, the current study positions Frog VLE as a target of ownership for which feelings of psychological ownership can be developed by the teachers who use it.

Research have linked and empirically found relationships between users’ participation (Barki, et al., 2008; Paré, Sicotte, & Jacques, 2006; Shen et al., 2013) and psychological ownership (Klesel et al., 2016; Smith, Grant, & Ramirez, 2014) on users’ beliefs towards technology. The logic is apparent as users participate and become involve in a system, the system become personally relevant and important to them, thereby inducing their feelings of ownership and buy-in for the particular system (Spears & Barki, 2010; Wu & Marakas, 2006). Transposing these relationships to the current context, the objective of the present study is to explore a model of users’ belief (perceived usefulness and perceived ease of use) that incorporates participation and psychological ownership. Although research has advanced the understanding of TAM variables, the rapid and continuous development of new technologies such as cloud-based VLE may open new directions of research that can enhance understanding of users’ beliefs.

2.2 Research Model

This study focused on the premise that teachers who participate in activities related to a VLE may be perceived to possess substantial influence on the VLE. This may elicit their feelings of psychological ownership for the VLE (POVLE) and influence their beliefs (PU and PEOU) towards the VLE. The study adapted the primary dimensions of overall responsibility, hands-on activities, and communication to measure participation (Barki et al., 2008; Paré et al., 2006). Examining these dimensions can greatly benefit any system, as participatory activities and assignments can be translated to actionable intervention beyond the initial phase of implementation (Venkatesh & Bala, 2008). With this, a research model is proposed in Figure 1.

![Figure 1. Research Model](image-url)
The following hypotheses underpin the study:

H1: Participation has significant influence on psychological ownership of VLE.
H2: Psychological ownership of VLE has significant influence on perceived usefulness.
H3: Psychological ownership of VLE has significant influence on perceived ease of use.

3. Methods

3.1 Participants

Cluster sampling was employed where 67 teachers (female = 46, male = 21) from three secondary schools responded to the study. This sample size fulfilled the criteria required as calculated by G*Power 3.1 software, with a statistical power of 80%, significance at 0.05 alpha level, and effect size ($f^2$) of 0.15 (Faul, Erdfelder, Buchner, & Lang, 2009). Teachers’ age ranged from 25 to 55, with a mean of 39 (SD = 8.55). Out of the 67 respondents, 62 of them had received training on the VLE. The average experience of using the system was 2.7 years (SD = 1.15), with usage experience ranging from one to five years. Only teachers who had experience using the system were included in the study, where respondents were filtered with a question asking whether they had prior experience in using the system. Permissions for research were given by various departments of the Ministry of Education, the school authorities, and the university where the researchers are based.

3.2 Measures

Participation is measured as a formative construct contributed by overall responsibility, hands-on activities, and communication (Barki et al., 2008). Formative measurement in Information System research is encouraged as it can help practitioners identify the specific dimensions for improvement (Mathieson, Peacock, & Chin, 2001). Hence, findings can provide information about the critical dimensions of participation. In contrast, the constructs of POVLE, PU and PEOU use reflective measurement (Barki et al., 2008; Wang & Wang, 2009). As both formative and reflective scales are used, common method bias will be reduced (Hair, Hult, Ringle, & Sarstedt, 2017).

Questionnaires were used to collect demographic data and to measure the constructs. Participation was measured with three dimensions of overall responsibility, hands-on activities, and communication. Indicators were adapted from Barki et al. (2008) to measure the extent of participatory activities with seven-point Likert scale ranging from “1 = never” to “7 = every time”. Psychological ownership of VLE was measured with items adapted from Barki et al. (2008), while user beliefs (perceived usefulness and perceived ease of use) were measured with items adapted from Wang and Wang (2009). Items were measured with a seven-point Likert scale with anchors ranging from “1 = strongly disagree” to “7 = strongly agree”. The instrument was reviewed by three subject matter experts for content suitability.

4. Findings

Partial Least Square Structural Equation Modelling holds an advantage in this study as it allows combination of reflective and formative measurement in the same model (Becker et al, 2012). Hence, data were analysed with SmartPLS 3.0 software (Ringle, Wende, & Becker, 2015).

4.1 Assessment of Reflective Measurement Model

The internal consistency of POVLE, PU and PEOU were verified with composite reliability of 0.886 to 0.915, and Cronbach’s Alpha values of 0.829 to 0.873 (Hair et al., 2017). Indicators which did not meet the loading threshold of .708 were removed (Hair et al., 2017), resulting the final indicators to load highest for their designated construct (Refer to Table 1). Convergent validity was verified with average variance extracted (AVE) of larger than 0.50 (Hair et al., 2017). In terms of discriminant validity, Table
2 displays the Fornell-Larcker’s (1981) criterion which were found that square roots of AVEs were greater than the correlations between the constructs and other constructs, denoting that these constructs are distinctively different from one another. This is further complemented in Table 3 with heterotrait-monotrait (HTMT) criterion which is more stringent that the Fornell-Larcker criterion (Henseler, Ringle, & Sarstedt, 2015) with values lower than HTMT.85 (Kline, 2011).

Table 1: Measurement model assessment for reflective model

<table>
<thead>
<tr>
<th>Construct</th>
<th>Indicators</th>
<th>Loadings</th>
<th>Composite Reliability</th>
<th>AVE</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychological ownership</td>
<td>PO1</td>
<td>0.732</td>
<td>0.902</td>
<td>0.699</td>
<td>0.855</td>
</tr>
<tr>
<td></td>
<td>PO2</td>
<td>0.791</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PO3</td>
<td>0.899</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PO4</td>
<td>0.909</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>PU1</td>
<td>0.919</td>
<td>0.915</td>
<td>0.732</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>PU2</td>
<td>0.704</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU3</td>
<td>0.918</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU4</td>
<td>0.868</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>PE1</td>
<td>0.823</td>
<td>0.886</td>
<td>0.667</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>PE2</td>
<td>0.817</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE3</td>
<td>0.761</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE4</td>
<td>0.846</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Fornell and Larcker criterion

<table>
<thead>
<tr>
<th>PEOU</th>
<th>POVLE</th>
<th>PU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.812</td>
<td>0.632</td>
<td>0.720</td>
</tr>
</tbody>
</table>

Table 3: HTMT criterion

<table>
<thead>
<tr>
<th>PEOU</th>
<th>POVLE</th>
<th>PU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>POVLE</td>
<td>PU</td>
</tr>
<tr>
<td></td>
<td>0.704</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.841</td>
</tr>
</tbody>
</table>

4.2 Assessment of Formative Measurement Construct

Participation is constructed as a formative variable contributed by three dimensions (overall responsibility, hands-on activities, communication). The summed score for each dimension was used as a formative indicator for participation (Barki et al., 2008). The construct was subject to redundancy analysis, assessment of multicollinearity between formative indicators, and assessment of significance and relevance of indicators (Hair et al., 2017). Convergent validity was tested with redundancy analysis with a global item of participation (Chin, 1998). As shown in Figure 2, the path coefficient linking the formative and reflective construct of participation was 0.798, demonstrating sufficient convergent validity with value of greater than 0.700 (Hair et al., 2017). This showed that participation is well represented by the three dimensions of overall responsibility, hands-on activities, and communication.

Figure 2. Redundancy Analysis for Assessing Convergent Validity
Collinearity between indicators may be a threat to the estimation of weights and statistical significance of the indicators as predictors of the formative constructs. Table 4 shows that variance inflation factor (VIF) values for each indicator are lower than 3.3, suggesting that the indicators measure different aspects of participation (Diamantopoulos & Siguaw, 2006). With this, bootstrapping of 5000 cases produced the outer weights of these indicators, which allows for the evaluation of their relative contribution to participation. Overall responsibility and hands-on activities were significantly related with participation (Table 4). Communication did not significantly contribute to participation, but potentially demonstrated absolute contribution with outer loading of more than 0.5 (0.675, p < .001) (Hair et al., 2017). Thus, communication is retained based on literature relevance to maintain the content validity of participation (Barki et al., 2008; Hartwick & Barki, 2001).

Table 4: Assessment of formative construct

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items</th>
<th>Convergent Validity</th>
<th>Weights</th>
<th>SE</th>
<th>VIF</th>
<th>t-value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation</td>
<td>Overall Responsibility</td>
<td>0.798</td>
<td>0.572</td>
<td>0.139</td>
<td>1.938</td>
<td>4.126**</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Hands-on Activities</td>
<td>0.467</td>
<td>0.16</td>
<td>2.248</td>
<td>2.927**</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>0.072</td>
<td>0.125</td>
<td>1.72</td>
<td>0.571</td>
<td>0.568</td>
<td></td>
</tr>
</tbody>
</table>

Note: ** >1.96, (two-tailed)

4.3 Assessment of Structural Model

As shown in the research model (Figure 1), each construct is only explained by one construct, hence, collinearity assessment for the structural model is not applicable in this study. Table 5 shows the results of path co-efficient assessment with bootstrapping of 5000 cases. The hypothesised relationships were found to be significant as demonstrated by their t-values, p values, and also without a zero being straddled in between the 95% confidence intervals. Participation explained 43.8% of POVLE, while POVLE explained 34.7% and 39.9% of PU and PEOU respectively. These $R^2$ values were above 0.26 which could be considered substantial (Cohen, 1988).

Table 5: Assessment of Hypothesised Relationships

<table>
<thead>
<tr>
<th>Hypothesis and Relationship</th>
<th>Direct Effect β</th>
<th>Standard Error</th>
<th>t-value</th>
<th>p value</th>
<th>Confidence Interval</th>
<th>Decision</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: UP → POVLE</td>
<td>0.661</td>
<td>0.073</td>
<td>9.023**</td>
<td>0.000</td>
<td>[0.468, 0.773]</td>
<td>Supported</td>
<td>0.438</td>
</tr>
<tr>
<td>H2: POVLE → PU</td>
<td>0.589</td>
<td>0.094</td>
<td>6.278**</td>
<td>0.000</td>
<td>[0.370, 0.739]</td>
<td>Supported</td>
<td>0.347</td>
</tr>
<tr>
<td>H3: POVLE → PEOU</td>
<td>0.632</td>
<td>0.084</td>
<td>7.516**</td>
<td>0.000</td>
<td>[0.425, 0.761]</td>
<td>Supported</td>
<td>0.399</td>
</tr>
</tbody>
</table>

Note: ** >1.645, (one-tailed)

5. Discussion

This study incorporated participation and psychological ownership in examining the TAM constructs of PU and PEOU. Participation was found to have significant effect on the psychological ownership of a VLE. Findings concurred with studies which assert that users feel they have greater influence on a system, thereby developing feelings of ownership for the particular system (Barki et al., 2008; Paré et al., 2006). The present study differs from the earlier ones on user participation, because it was operationalised as the extent of activities and behaviours teachers perform in the cloud VLE, as they tailor the applications and programs which run through the cloud for teaching and learning.

Findings showed that overall responsibility and hands-on activities with the exception of communication have significant contribution in forming participation. Teachers’ overall responsibility in the VLE include activities and assignment that entails their leadership and accountability for the system (Barki & Hartwick, 1994; Ju et al., 2016), while hands-on activities involve specific physical
design and implementation tasks performed by users (Barki et al., 2008; Hartwick & Barki, 2001). The main activities in a VLE include developing its virtual environment by building learning contents, and cloud-based VLEs allow high degree of versatility to do so, facilitated by seamless access to resources (Park & Ryoo, 2013; Stantchev et al., 2014). Such flexibility in tailoring and customizing may elicit teachers’ sense of ownership for the VLE, as they have to invest their time and intellect when carrying out activities in it and the resultant content embodies their desired outcome. As such, additional hands-on activities that lead to more responsibility and accountability can also be incorporated to develop teachers’ feelings of ownership.

Frog VLE as a target of psychological ownership was found to have significant effect on PEOU, and to a lesser extent, on PU. Indeed, this study implies the value of psychological ownership as a driver for these salient beliefs in technology acceptance and adoption. Interestingly, POVLE asserts greater effect on PEOU than PU. This suggests that psychological ownership may be more important to users’ belief about the ease of operating a cloud based VLE. Such VLE differs from traditional LMS because they are innovative learning systems that offer greater complex functionalities in terms of course management features and learning activities (Shiau & Chau, 2016). Psychological ownership may be useful in overcoming teachers’ perception of the complexity of such systems, and can enhance its perceived usefulness.

5.1 Conclusion and Future Investigations

This study explored the role of participation and psychological ownership on user beliefs about the cloud based VLE. Three relationships were hypothesised, and results concurred with previous studies which was carried out in previous information technology contexts. The hypothesised relationship suggested that the extent of teachers’ participation in the VLE can elicit their feelings of ownership towards the VLE. These feelings of psychological ownership for the VLE were found to have significant effect on teachers’ perceived usefulness and perceived ease of use of the VLE. As POVLE is developed after having some experience using the system, this construct can be useful for examining post-adoption behaviours of existing systems.

In spite of the magnitude of the influence found from the hypothesized relationships, the current study has limitations which need to be addressed in future investigations. Only two main TAM variables of PU and PEOU were examined. Future investigations may incorporate these constructs to investigate the TAM in its totality, to enhance and test the predictive value of the model. In addition, even though the quantitative evidence obtained demonstrated defensible validity and reliability, future research which include interviews can further strengthen the validity of the findings. In short, this study provided evidence for future investigations to explore the construct of participation, psychological ownership, and user beliefs in other educational technologies.

Acknowledgement

This paper is based on a research funded by Universiti Tunku Abdul Rahman Research Fund. The authors would like to express their gratitude to the university in sponsoring the research.

References


A Learning Support Method to Raise Awareness of the Knowledge-to-Action Gap in Information Ethics

Koji TANAKA*, Honomi MIWAb, Mitsuru IKEDAc & Masahiro HORIc

SaSchool of Knowledge Science, Japan Advanced Institute of Science and Technology, Japan
bGraduate School of Informatics, Kansai University, Japan
*cjtanaka@jaist.ac.jp

Abstract: If learners cannot form an intention to take an appropriate action regarding the knowledge of information ethics, it cannot be said that the learners acquired working knowledge of ethical principles. It is necessary to conduct a learning activity so that they can realize the difference between behaviors in a realm of learning/training situation and those in real-life situations. In the present study, we propose a learning support method to raise self-awareness of the inconsistency between knowledge and intention in information ethics. On the basis of the proposed method we develop an education program, and conduct the program for first-year university students. The program consists of two phases. In the first phase, students answered paper-and-pencil tests consisting of two tasks: behavior selection task to confirm the knowledge-to-action gap, and behavior evaluation task to ensure factors contributing to the gap. In the second phase, students were shown the graph of experimental results as feedback. As results of quantitative and qualitative analysis on questionnaire for evaluating learning activities, the program allowed students to capture the gap as their own matters and to recognize the activities as learning opportunity on information ethics. In addition, the program enhanced motivation of the participants toward learning information ethical behavior.

Keywords: knowledge-to-action gap, information ethics, learning support method

1. Introduction

As information technology goes through rapid advancement, each individual is increasingly required to have a maturity of information ethics. Ministry of Education, Japan (2009, p.37) defines Information ethics as “the thinking and attitude that serve as the foundation for conducting appropriate activities in information society.” As implied by this definition, one cannot say he or she has acquired working knowledge about information ethics simply by possessing knowledge on the matter unless behavioral intention to actualize such knowledge into action has formed. For this reason, it is crucial to develop teaching methods and materials designed to foster such an attitude toward actualizing knowledge of information ethics into action. Tamada and Matsuda (2000, 2004) advocated three kinds of knowledge play important roles in the education of information ethics. The three kinds of knowledge are: “knowledge on information technology” that is the characteristics and technological limitations of an information-based society, “knowledge on moral norms” that are norms to be adhered, and “knowledge on logical judgement” that is how to make appropriate decisions using the previous two kinds of knowledge and compare them to meet various value standards. In this study, we assume that having learners themselves recognize their own mental process that leads behaviors of going against information ethics (hereafter, “unethical behavior”) will aid in fostering the mindset for making appropriate judgments.

For learners to recognize their mental process, it is necessary to conduct a learning activity in which they realize the difference between behaviors they take in a realm of learning/training situation and behaviors they take in real-life situations. Therefore, this study proposes a learning support method to enhance learners’ motivation through self-awareness of one’s mental process when taking unethical behavior. The proposed method consists of two learning activities. The first activity is to participate in
questionnaire survey of information ethical behavior conducted as a cognitive psychological experiment (Tanaka, Ikeda, & Hori, 2016) to actualize the difference between knowledge as general principles and behavioral intention they may have in an actual situation. The second activity is to get feedback on the questionnaire results depicted as graph, thereby quantifying the difference between knowledge and behavioral intention.

In the next section, a brief overview of the proposed method is introduced. Section 3 explains an education program based on the proposed method and the way how to put the program into practice for a class of university students. Section 4 presents the result of the learning evaluation questionnaires in the class, and finally concluded with remarks on the further study.

2. Learning Support Method for Recognizing One’s Mental Process

Human mental processes have been studied in the field of cognitive psychology and cognitive science. Research on complex cognitive systems or psychological phenomena often employ model-based approaches (Fum, Del, & Stocco, 2007). Miwa et al. viewed model-based approaches not only as research tools for understanding internal processing but also as educational tools for deepening learner’s understanding of his/her own cognitive process (Miwa & Terai, 2016; Miwa, Terai, & Shibayama, 2016; Saito, et al., 2015). From such a point of view, they developed a model-based learning environment (DoCoPro) in which the learners construct their cognitive model. In DoCoPro, students learn how to automate the process of solving addition and subtraction problems by using rule-based descriptions of computer programs. In addition, Hulshof et al. demonstrated that subjective experience of a psychological phenomenon and reflection on such experience is effective for a better understanding of a psychological phenomenon (Hulshof, Eysink, Loyens, & De Jong, 2005; Hulshof, & Eysink, 2006). They developed a learning environment (ZAPs), which allows learners to experience a psychological phenomenon through participation into a psychological experiment, and facilitate understanding of the phenomenon by the feedback of experimental results. In ZAPs, students learn scientifically supported psychological phenomena by means of psychological experiments such as an illusion and a mental rotation.

An important idea behind these learning environments is to provide understanding of the mental process by having the learners participate in the actualization of their own mental process, so that they can recognize the process by themselves. In the education program of this study, we adopted a similar approach that includes participation of students in a psychological experiment followed by the feedback of experimental results. However, the learning objective of this study is different from the previous studies. The objective of this study is to make learners become self-aware of real-life issues caused by a mental process and enhance motivation toward learning through the recognition of one’s mental process (Figure 1). In the first phase of this education program, a behavior selection task was given as cognitive psychology experiment (Tanaka, Sonoda, Ikeda, and Hori, 2016a, 2016b) to actualize inconsistency between knowledge and behavioral intention in information ethics. This phase was designed to enhance readiness toward learning by promoting self-awareness of the inconsistency. In the second phase, after debriefing session of the behavior selection task, the learners were presented
with a graph showing the inconsistency between knowledge and behavioral intention. Note that the result quantified in the graph includes responses of the learners participated in the experiment in the first phase. The inconsistency revealed in the graph may not necessarily be acceptable for the learners. For example, learners may doubt that “it's not that simple” and that “there could be some other reasons”. Encountering such unconvinced situation can be momentum for the recognition of inconsistency underlying the learners’ mental process, and that will lead to their motivation for understanding the process more deeply. In contrast, in the cases of DoCoPro and Zaps, learners do not cast doubt about the results because their objective of learning is to absorb erudite knowledge.

3. Education Program

In this study, situations that require information ethical behaviors were sorted out into 20 items (Table 1) sourced from the six textbooks and five supplementary readings, as well as from the Metropolitan Police Department’s website. The education program was carried out in the spring semester of 2016 in a class of Introduction to Informatics offered for the first-year university students as an obligatory subject. An information ethics behavioral task (hereafter, “behavioral task”) was conducted in 25 minutes during the first lesson. Next week in the second lesson, task debriefing was conducted and the results of the task were presented in a graph (totally in 40 minutes).

3.1 Information Ethics Behavioral Task

In the behavioral task, the students were asked to engage in two tasks for behavior selection and evaluation. In this behavioral task, the students were placed in a situation in which they had to select an action taking information ethics into account. Our intention here was that the students involved in this task would realize the reality wherein one does not always behave in accordance with appropriate norm, even if they have the ability to comply with. This provides students an opportunity afterwards to think about making them conform to ethical behavior.

3.1.1 Information Ethics Behavior Selection Task

In the information ethics behavior selection task (hereafter, “selection task”), the students were presented with texts explaining the situation setting of each task where information ethical behavior is needed. The students were then presented with the conflicting options of an ethical and unethical behavior. The students responded to the knowledge task (“Which behavior is appropriate in principle when using information technology?”) and the intention task (“Which behavior would you yourself choose?”). The situation setting text and subsequent two-alternative questions were prepared based on the descriptions given in textbooks and supplementary readings. In the case of texting while walking, for example, the situation setting text was given as “You receive a message on your smartphone (or

<table>
<thead>
<tr>
<th>Ethics in information society</th>
<th>Information security</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chainmail forwarding</td>
<td>Without renewing anti-virus software</td>
</tr>
<tr>
<td>Identification of individuals on SNS</td>
<td>Without preventing data loss</td>
</tr>
<tr>
<td>Phones in priority seating areas</td>
<td>Password recycling</td>
</tr>
<tr>
<td>Digital theft</td>
<td>Password storage</td>
</tr>
<tr>
<td>Texting while walking</td>
<td>Without confirming the URL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Understanding and observing with the law</th>
<th>Wisdom for safety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright infringement</td>
<td>Without confirming the reliability of information</td>
</tr>
<tr>
<td>Portrait right infringement</td>
<td>Paying false bills</td>
</tr>
<tr>
<td>Fraudulent access</td>
<td>Replying to unknown senders</td>
</tr>
<tr>
<td>Illegal downloading</td>
<td>Without confirming use of personal data</td>
</tr>
<tr>
<td>Electromagnetic records by illegal command</td>
<td>Real-time posting</td>
</tr>
</tbody>
</table>
mobile phone) while you are out shopping in town.” An option was “You step out of the way and stop to check your mail” (ethical behavior), and another option was “You stay out of the way but continue walking as you check your mail” (unethical behavior). Other examples of tasks are found in the paper (Tanaka, Sonoda, Ikeda, and Hori, 2016b).

In the inconsistency quantification method, by comparing the student’s responses for the knowledge and intention tasks, it is determined whether an inconsistent judgement occurs or not. Results of the selection task are summarized as follows: for all items, other than the forwarding of chain mails and payment to fictitious claims, the correct response rate in the intention tasks was significantly lower than that in the knowledge tasks, which confirmed an incongruity between knowledge and behavioral intention (Tanaka, Miwa, Ikeda, & Hori, 2016).

3.1.2 Information ethics behavior evaluation task

In the information ethics behavior evaluation task (hereafter, “evaluation task”), following the situation setting text for each ethical behavior, either an ethical or unethical behavior was presented to the students. In this task, we considered the three factors: attitude toward the presented action, subjective norms, and students’ perceived behavioral control, which were said to influence the behavioral intention according to the theory of planned behavior (Ajzen, 1991). The students were then asked to respond to the question: to what extent do the factors apply to their mindset on a seven-point evaluation scale. Each scale was quoted from Ajzen’s paper (Beck & Ajzen, 1988) and translated to Japanese.

Results of the evaluation task are summarized as follows: for all the factors, the evaluation rates were significantly different between students selected ethical behavior in the intention tasks and students selected unethical behavior (Tanaka, Miwa, Ikeda, & Hori, 2016).

3.2 Presentation of the graph of information ethics behavioral task results

Based on the results of selection tasks in which 448 first-year university students (who attended the first lesson) responded, a graph that shows the inconsistency between one’s knowledge and behavioral intention (hereafter, “inconsistency graph”) was created. As shown in Figure 2 (a), the inconsistency graph was a column graph where the y-axis indicated the percentage of students who selected an ethical action in the knowledge and intention tasks, while the x-axis showed the labels for the 20 items. The students were explained that the differences in the height between the two tasks revealed inconsistency.

Before presenting the graph, a debriefing was conducted to explain the situation setting for the selection task and ethical/unethical behaviors given in each setting. The inconsistency graph was then presented on the screen in the classroom. All the materials used for these explanations are prepared as a booklet, and copies of the booklet were distributed to all the attendees.

Next, based on the results of the evaluation task, another graph with the three factors of unethical behavior (hereafter, “factor graph”) was displayed on the screen and reproduced in the distributed booklet. Before presenting the factor graph, a briefing was conducted using the screen and the booklet, with the explanations on the question texts (factors) in the evaluation task. As shown in Figure 2 (b), the factor graph was a column graph where the y-axis indicated the mean evaluation value for the factors, while the x-axis showed the behavior selected in the selection task as well as the behavior evaluated in the evaluation task. The students were explained that how the evaluation of factors was different between the students selected the unethical behavior in the intention task and the students selected the ethical behavior.

4. Analysis of Responses from the Learning Evaluation Survey

After the presentation of the inconsistency graph, a survey was conducted to evaluate the learning activities by two kinds of questions on a seven-point scale (1: Strongly disagree, 7: Strongly agree) and free description of the reason for selecting the scale. The first question was to confirm learning opportunity (hereafter, “opportunity question”) by asking “Do you feel that this served as an opportunity to think about information ethics?” and the second question to confirm learning motivation (“hereafter, “motivation question”) by asking “Do you think this program promoted your motivation for
learning information ethics?" In addition, after the opportunity and motivation questions, a survey was conducted to evaluate the learning activities in the evaluation task by question on a seven-point scale (1: Strongly uninterested, 7: Strongly interested). The question was to confirm the degree of interest in three factors (hereafter, “interest question”) by asking “Do you have an interest in each factor of information ethical behavior?” In these learning evaluation surveys, a total of 394 first-year students, who had responded to the behavioral task in the first lesson, were included as analysis subjects.

### 4.1 Seven-point Scale Evaluation

To examine the effects of engagement in the selection task and reading the inconsistency graph, the seven-point evaluation values in the learning opportunity and motivation questions were divided into two categories: positive evaluation (P-value), and non-positive evaluation (NP-value). The P-value is the cases when the evaluation was higher than the neutral, namely, the value is either 5, 6 or 7, while the NP-value is the cases when the value is either 1, 2, 3 or 4. A binomial test was performed on the number of students whose responses were P-value or NP-value. Considering the number of values allocated to each evaluation category, the test ratio was set as 3:4. For both opportunity and motivation questions, the result demonstrated that the number of students selected P-value was significantly higher than those who selected NP-value (Table 2). From this result, engaging in the selection task and reading the inconsistency graph could be helpful for providing learning opportunities of information ethical behavior and enhancing learning motivation.

### Table 2: Number of students in each evaluation.

<table>
<thead>
<tr>
<th></th>
<th>P-value</th>
<th>NP-value</th>
<th>Binomial test [95%CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral selection in the selection task</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opportunity Question</td>
<td>270</td>
<td>124</td>
<td>(p &lt; .001)** [.64,.73]</td>
</tr>
<tr>
<td>Motivation Question</td>
<td>196</td>
<td>198</td>
<td>(p = .003)** [.45,.55]</td>
</tr>
<tr>
<td>Expected Frequency</td>
<td>168.86</td>
<td>225.14</td>
<td></td>
</tr>
<tr>
<td>Seeing of the inconsistency graph</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opportunity Question</td>
<td>266</td>
<td>128</td>
<td>(p &lt; .001)** [.63,.21]</td>
</tr>
<tr>
<td>Motivation Question</td>
<td>213</td>
<td>181</td>
<td>(p &lt; .001)** [.49,.59]</td>
</tr>
<tr>
<td>Expected Frequency</td>
<td>168.86</td>
<td>225.14</td>
<td></td>
</tr>
</tbody>
</table>

Note. \(N = 394\). CI = confidence interval. ** \(p < .01\).

### Table 3: Coding rules.

<table>
<thead>
<tr>
<th>Code name</th>
<th>Main words used for coding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Myself and others</strong></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>Applies to myself myself, personal, self-awareness, empathy</td>
</tr>
<tr>
<td>A2</td>
<td>One’s surrounding real-life, everyday, usually, familiar</td>
</tr>
<tr>
<td>A3</td>
<td>Family and Friends family, friend</td>
</tr>
<tr>
<td>A4</td>
<td>Others others, neighborhood, opinions around me, behavior around me</td>
</tr>
<tr>
<td><strong>Knowledge-to-action gap</strong></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Inconsistency inconsistency, incongruity, discrepancy, conflict in behavior</td>
</tr>
<tr>
<td>B2</td>
<td>Knowledge common sense</td>
</tr>
<tr>
<td>B3</td>
<td>Behavior behave, action, [before or after the word do/take] act or behavior</td>
</tr>
<tr>
<td><strong>Learning opportunity</strong></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Interest interest, concern, doubt, wonder, why</td>
</tr>
<tr>
<td>C2</td>
<td>Unexpectedness surprising, unexpectedly, shock, impact</td>
</tr>
<tr>
<td>C3</td>
<td>Realization realization, recognition, confirmation, find out, actual feeling</td>
</tr>
<tr>
<td><strong>Learning motivation</strong></td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>Consciousness consen, keep in mind, [before or after the word eye] turn</td>
</tr>
<tr>
<td>D2</td>
<td>Reviewing review, reflect, re-think, change one's mind</td>
</tr>
<tr>
<td>D3</td>
<td>Learning learn, be taught</td>
</tr>
</tbody>
</table>
Furthermore, to examine the effects of engagement in the evaluation task, the seven-point evaluation values in the interest question were divided into two categories: interested evaluation (I-value) for point values 5, 6 or 7; and uninterested evaluation (UI-value) for point values 1, 2, 3 or 4. A binomial test was conducted on the number of students whose responses were I-value or UI-value. Considering the number of values allocated to each evaluation category, the test ratio was set as 3:4. For each of the three factors, the result demonstrated that the number of students selected I-value was significantly higher than those who selected UI-value (Attitude: 187 vs. 207, \( p = .04 \), Subjective norms: 199 vs. 195, \( p = .001 \), Perceived behavioral control: 209 vs. 185, \( p < .001 \)). From these results, engaging in the evaluation task could be helpful for generating interest in the factors of information ethical behavior.

4.2 Quantitative Text Analysis

To clarify the overall tendency of the responses filled in the freewriting spaces, a weighted text analysis was conducted with a software for qualitative content analysis called KH Coder (Higuchi, 2014). Table 3 shows the list of coding rules and the key terms for coding. For example, key terms either ‘myself’ or ‘(became) aware’ is found in a paragraph, a code “apply to myself” is assigned to the paragraph.

In this quantitative text analysis, it is examined what kind of learning opportunities is provided with students as the result of each learning activity. A nominal variable here indicates the two learning activities: the behavioral selection in the selection task and the reading of the inconsistency graph. We performed a cross tabulation and \( \chi^2 \) test to confirm the difference between the number of learning activity codes appeared in the learning activities (Table 4). The result showed there was no significant difference in the occurrence of “applies to myself” code (A1). Concerning “one’s surrounding” (A2) and “(became) aware” (D2), a significant difference was confirmed in the number of occurrences, where the occurrence in comments for the behavioral selection was higher than that for the reading of the graph. Concerning the “others” (A4), “inconsistency” (B1), “interest” (C1), and “unexpectedness” (C2), a significant difference was found, where the occurrence in the graph reading was higher than that in the behavioral selection. Taking account of the terms such as “one’s surrounding” and “(became) aware” appeared in the comments on the behavioral selection, it can be inferred that the information ethics behavior selected for questioning by the students provided them with a good opportunity of taking a look at the behavior of people around. On the other hand, the terms such as “others,” “inconsistency,” “interest,” and “unexpectedness” were found in the comments on the graph reading. This result suggests

### Table 4: Cross tabulation and \( \chi^2 \) test in confirmation question for learning opportunity.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>Number of case</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>76</td>
<td>44</td>
<td>18</td>
<td>14</td>
<td>6</td>
<td>18</td>
<td>25</td>
<td>4</td>
<td>111</td>
<td>17</td>
<td>35</td>
<td>8</td>
<td>344</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(22.09%)</td>
<td>(12.79%)</td>
<td>(5.23%)</td>
<td>(4.07%)</td>
<td>(1.74%)</td>
<td>(5.23%)</td>
<td>(7.27%)</td>
<td>(1.16%)</td>
<td>(32.27%)</td>
<td>(4.94%)</td>
<td>(10.17%)</td>
<td>(2.33%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>65</td>
<td>9</td>
<td>0</td>
<td>60</td>
<td>30</td>
<td>3</td>
<td>11</td>
<td>40</td>
<td>15</td>
<td>114</td>
<td>12</td>
<td>6</td>
<td>318</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(20.44%)</td>
<td>(2.83%)</td>
<td>(0.00%)</td>
<td>(18.87%)</td>
<td>(9.43%)</td>
<td>(0.94%)</td>
<td>(3.46%)</td>
<td>(12.58%)</td>
<td>(4.72%)</td>
<td>(35.85%)</td>
<td>(3.77%)</td>
<td>(5.03%)</td>
<td>(1.89%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>141</td>
<td>53</td>
<td>1</td>
<td>78</td>
<td>44</td>
<td>9</td>
<td>29</td>
<td>65</td>
<td>19</td>
<td>225</td>
<td>51</td>
<td>14</td>
<td>662</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(21.30%)</td>
<td>(8.01%)</td>
<td>(0.15%)</td>
<td>(11.78%)</td>
<td>(6.65%)</td>
<td>(1.36%)</td>
<td>(4.38%)</td>
<td>(9.82%)</td>
<td>(2.87%)</td>
<td>(33.99%)</td>
<td>(4.38%)</td>
<td>(7.70%)</td>
<td>(2.11%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \chi^2 )-test</td>
<td>0.18</td>
<td>20.93**</td>
<td>0.00</td>
<td>28.26**</td>
<td>6.82**</td>
<td>0.31</td>
<td>0.85</td>
<td>4.68*</td>
<td>6.27*</td>
<td>0.79</td>
<td>0.30</td>
<td>5.45*</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5: Cross tabulation and \( \chi^2 \) test in confirmation question for learning motivation.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>Number of case</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>56</td>
<td>10</td>
<td>2</td>
<td>19</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>54</td>
<td>5</td>
<td>101</td>
<td>1</td>
<td>6</td>
<td>65</td>
<td>322</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(17.39%)</td>
<td>(3.11%)</td>
<td>(0.62%)</td>
<td>(5.90%)</td>
<td>(3.11%)</td>
<td>(2.17%)</td>
<td>(2.80%)</td>
<td>(16.77%)</td>
<td>(1.55%)</td>
<td>(31.37%)</td>
<td>(0.31%)</td>
<td>(1.86%)</td>
<td>(20.19%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>44</td>
<td>5</td>
<td>0</td>
<td>36</td>
<td>29</td>
<td>5</td>
<td>10</td>
<td>61</td>
<td>6</td>
<td>111</td>
<td>11</td>
<td>6</td>
<td>59</td>
<td>313</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(14.66%)</td>
<td>(1.60%)</td>
<td>(0.00%)</td>
<td>(11.50%)</td>
<td>(9.27%)</td>
<td>(1.60%)</td>
<td>(3.19%)</td>
<td>(19.49%)</td>
<td>(1.92%)</td>
<td>(33.46%)</td>
<td>(3.51%)</td>
<td>(1.92%)</td>
<td>(18.85%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>15</td>
<td>2</td>
<td>55</td>
<td>39</td>
<td>12</td>
<td>19</td>
<td>115</td>
<td>11</td>
<td>212</td>
<td>12</td>
<td>12</td>
<td>124</td>
<td>635</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15.75%)</td>
<td>(2.36%)</td>
<td>(0.31%)</td>
<td>(8.66%)</td>
<td>(6.14%)</td>
<td>(1.39%)</td>
<td>(2.99%)</td>
<td>(18.11%)</td>
<td>(1.72%)</td>
<td>(33.39%)</td>
<td>(1.89%)</td>
<td>(18.93%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \chi^2 )-test</td>
<td>1.09</td>
<td>0.98</td>
<td>0.474</td>
<td>5.61*</td>
<td>9.41**</td>
<td>0.06</td>
<td>0.00</td>
<td>0.62</td>
<td>0.00</td>
<td>1.02</td>
<td>7.14**</td>
<td>0.00</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. \( N = 394 \). BS = behavioral selection in the selection task, RG = reading of the inconsistency graph. A1 = applies to myself, A2 = one’s surrounding, A3 = family and friends, A4 = others, B1 = inconsistency, B2 = knowledge, B3 = behavior, C1 = interest, C2 = unexpectedness, C3 = realization, D1 = consciousness, D2 = reviewing, D3 = learning. * \( p < .05 \), ** \( p < .01 \).
that receiving a debriefing on the selection task and the inconsistency graph will provide opportunities for paying attention to the belief of classmates regarding the gap between knowledge and behavioral intention.

Next, it is examined whether students’ motivation for learning is improved as the result of each learning activity. To confirm the difference in the number of occurrences of the codes found in the survey comments, a cross tabulation and $X^2$ test were conducted (Table 5). The result showed there was no significant difference in the occurrence of “applies to myself” code (A1). Concerning “others” (A4), “inconsistency” (B1) and “consciousness” (D1), a significant difference was found, where the occurrence in comments for the graph reading was higher than that for the behavioral selection. Since such terms as “others,” “inconsistency” and “consciousness” were written down after the graph reading, it is inferred that the debriefing on the selection task and inconsistency graph promoted learning motivation of the students as well as consciousness toward the inconsistency between other people’s knowledge and behavioral intention, conceived by their classmates. These results demonstrated that the two learning activities in this education program contributed to achieving the learning objectives.

Furthermore, to confirm the difference in the number of occurrences of codes between the evaluator group for each question, the nominal variable (evaluator group) corresponding to positive and non-positive responses were prepared, and a cross tabulation and $X^2$ test were conducted. As the result, the number of occurrences of the code “applies to myself” was higher among the P-evaluator group than the NP-evaluator group in all questions of the learning activity (In behavioral selection, opportunity question: $X^2 = 10.20, p < .01$, motivation question: $X^2 = 19.79, p < .01$. In graph reading, opportunity question: $X^2 = 5.16, p < .05$, motivation question: $X^2 = 11.84, p < .01$). In the behavioral selection, the number of occurrences of the code “realization” was higher in the NP-evaluator group than the P-evaluator group in the opportunity question ($X^2=4.84, p < .05$). The similar trend was observed in cases of “one’s surrounding ($X^2= 7.48, p < .01$)” for learning eagerness in behavioral selection, “inconsistency ($X^2= 4.14, p < .05$)” in the learning impetus in the comments on the graph, and “others ($X^2= 8.25, p < .01$),” “inconsistency ($X^2= 7.61 p < .01$),” and “realization ($X^2= 5.02, p < .01$)” in the learning eagerness in the comments on the graph. No significant difference was confirmed between evaluators in other codes.

Consequently, the implications from the above results were summarized as follows. Learners who positively evaluated the education program became aware of the inconsistency between knowledge and behavioral intention of their own due to the engagement in the behavior selection task. Moreover, the positive learners were aware of inconsistent beliefs of others as the result of engaging in not only the selection task but the reading of the inconsistency graph.

4.3 Descriptive Contents of Freewriting Spaces

The two kinds of descriptions written by learners are connected with deep thinking in a learning activity (Suda, 2017). The one is description of the change of consciousness that learner links past and future self with learning contents. Another is description of the reconfiguration of knowledge that learner doubts and/or considers about learning contents. In what follows, we focus on as the above two kinds descriptions and interpret the response of the freewriting responses at a semantic level, a content analysis of the descriptions was conducted in terms of the learning opportunity and motivation.

4.3.1 Opportunity for Learning Information Ethics

It was observed that self-awareness of unethical behavior was manifested from descriptions. Examples of such descriptions are: “It became commonplace for me to deal with information, and I was doing things that aren’t correct as if it’s something normal. This learning made me realize this fact once again” (opportunity question about behavioral selection), and “Because there were many things that I did unconsciously, this learning activity made me check my behaviors once again” (opportunity question about behavioral selection).

There were also descriptions showing the self-awareness of inconsistency between knowledge and behaviors, such as “This learning activity served as an opportunity to find the unethical behaviors [I know shouldn't do it but I do because I think it’s something that’s not that important] which took by myself on a regular basis” (opportunity question about behavioral selection), and “Although I had
knowledge on information ethics, it made me realize that it wasn’t seen in my behaviors under various circumstances” (opportunity question about behavioral selection).

Furthermore, there were also descriptions comparing oneself with others as to the inconsistency between knowledge and behavior, such as “When I actually looked at the graph, there were many people who couldn’t behave despite having knowledge like me. It made me to re-consider about information ethics” (opportunity question about the graph reading), and “There was a little thing within me that was not able to make the correct selection. The people around me seemed to be unable to do so more than me. So, I felt that I had to behave more properly” (opportunity question about the graph reading). These descriptions suggested that the education program in this study worked for creating awareness of the necessity to learn of information ethics.

4.3.2 Motivating Information Ethics Learning

There were descriptions of concerns about how to learn the ways of resolving the inconsistency between knowledge and behavior, such as “Information ethics is something that everyone has general knowledge of --almost without a doubt-- so, conversely, I would want to learn the means for solving this inconsistency” (motivation question about the graph reading), and “I thought that rather than learning about ‘the importance of information ethics,’ I would like to know more about preventative measures not to take unethical behavior” (motivation question about behavioral selection).

The other descriptions mentioned the factors that could influence the occurrence of inconsistency between knowledge and behavior, such as “I would like to think deeply as to why there were several people who knew about information ethics as common sense and as morality but could not put into behavior” (motivation question about the graph reading) or “This learning activity made me curious as to its reason, because I could confirm that what kinds of items was inconsistent and how obvious the differences were at a glance” (motivation question about the graph reading).

Furthermore, some students thought unethical behavior covered in the education program was limited, and motivated to know more in depth about information ethical behavior; for example, “Simple questions as seen in the previous questionnaire made me understand which behavior is correct and which one is incorrect. However, it made me want to learn because I don’t think I could have answered correctly at all, if these topics had gone deeper” (motivation question about behavioral selection), and “I think there are more inconsistencies between knowledge and behavior other than the ones seen here. It aroused my interests more.” (motivation question about behavioral selection). From these descriptions, the learning activities in this education program can be anticipated to promote students’ motivation toward learning information ethics.

4.3.3 Dissatisfaction with Learning Activity

On the other hand, there were students who did not perceive this education program positively. Some students commented that they had already known appropriate behavior as knowledge: “I have already learned most of the knowledge in high school” (opportunity question about graph reading), and “I already know a lot of matters. I don’t particularly feel the need for further learning about information ethics” (opportunity question about behavioral selection).

There were also accounts that attributed the inconsistency as a known fact, such as “Indeed, I think it served as an opportunity to think about information ethics. However, I suppose the knowledge and behaviors are evident, and don’t think it served as a major impetus” (opportunity question about behavioral selection), and “The result of analysis presented with graph and explained point by point. However, it seemed to be obvious, and the analysis did not affect me for the further learning” (opportunity question about graph reading).

From these descriptions, students with unfavorable opinions perceived the objective of this education program as a learning activity simply to acquire the knowledge and facts without self-reflection on the behavior.
5. Concluding Remarks

Based on the analysis of responses to the learning evaluation survey, it was observed that most students reviewed information ethics behavior as matters of their own concern, by having engaged in the selection task whereby selecting behavior they would take. As the result of receiving a debriefing on the task and reviewing the inconsistency graph, it was confirmed that students had interest in the inconsistency between knowledge and behaviors, becoming aware of inconsistent beliefs of others. Furthermore, it is expected the students attain the motivation to learn more about the solution and factors, considering the causes of inconsistency.

On the other hand, the current limitation of the proposed method comes from the fact that some students perceived an objective of this education program as simply acquiring the knowledge and well-known facts in information ethics. When experimental result is ensured with external validity, which means the result of the study can be generalized to other situations, implications of the result can be explained in general terms taking the dependence on individual situations into account. Therefore, to teach the significance of the experimental result as behavioral principles, learners have to understand the principles in relation to their own experiences, not merely in general terms. In this sense, it is important to note that this education program allowed students to be aware of the inconsistent beliefs of not only their own but also others. Such awareness is a key to further study of this learning support method, which is conducted in part as a cognitive psychological experiment and intends to facilitate understanding of mental process behind the ethical behaviors.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26560133 and 16K12782.

References


An Empirical Study on the Influencing Factors of ICT Application: A Large Scale Survey for Middle and Primary Schools in China

Wei WANG* & Chun LUb*

*a National Engineering Research Center for E-Learning, Central China Normal University, China
b The Education Ministry Strategic Research Base of Educational informatization (Central China), Central China Normal University, China

*luchun@mail.ccnu.edu.cn

Abstract: ICT application is the core driving force for optimizing the development of educational information system. In this paper, to explore the different influencing factors of ICT application in middle and primary schools, the following works have been done: Firstly, an exploratory factor analysis is used to classify ICT application into three dimensions. Secondly, a series of linear regression models are used to test whether ICT policy influences ICT application. Finally, by analyzing the different influencing effects in urban and rural areas, combined with the actual situation, the paper made a comparative analysis for influencing factors of ICT application between the urban and rural middle and primary schools. In this paper, we found that: 1. ICT application could be investigated from three aspects: digital educational resources, online learning and educational environment. 2. There was a certain degree of correlation between ICT application and ICT policy in schools of all regions, that is to say ICT policy can influence ICT application. 3. Because ICT application in urban and rural areas were at different stages, different school policies also affected them differently. In the conclusion, influencing factors on ICT application for schools in urban and rural areas were different, then we put forward some suggestions for ICT policy at the school-level respectively. We hope the way China coped with can provide the developing countries even underdeveloped areas, such as the African region with some feasible ideas.

Keywords: ICT application; influencing factors; middle and primary schools; exploratory factor analysis; linear regression; comparative analysis

1. Introduction

With ICT widely applied in the field of education, it has greatly promoted the innovation and development of educational philosophy, teaching mode and pedagogical methodology, so that the development of ICT in education is gradually changed from construction-driven to application-driven. In recent years, ICT application in education has received wide attentions. A study showed that the use of ICT in education is very important from the perspective of teachers (Trucano, 2005). In 2009, Education, Audiovisual and Cultural Executive Agency (EAECA) published an official report, which had established the indicators’ framework of ICT application for middle and primary schools including the use intension from teachers, “Computer use” and “Internet use” from students (Pelgrum, 2009). British Educational Communications and Technology Agency (BECTA) published the self-review framework (SRF), which pointed out the item about ‘ICT application in Curriculum’ to evaluate the development and application of ICT in schools (BECTA, 2011).

In the literature, many studies on the factors influencing ICT application in education have been focused. According to the Technology Acceptance Model (TAM) used to predict application of ICT by agricultural students, authors explored direct and indirect effects on ICT application, and found that ‘skill’, ‘support’, and ‘facilities’ were the three factors that influenced ICT application by agricultural students (Pouratashi & Rezvanfar, 2010). Moreover, ‘limitations of hardware’, ‘lack of in-service training’ and ‘lack of technical support’ were the most important barriers for primary school teachers in the ICT integration (Goktas, 2013), which could be inferred that the improvement of hardware,
teachers’ training and technical support were favor in ICT application. Research has shown that the success of the implementation of ICT is not dependent on the availability or absence of one individual factor, but is determined through a dynamic process involving a set of interrelated factors (Afshari, Bakar, Luan, Samah, & Fooi, 2009). Selecting some measurable indicators, Chun Lu conducted an empirical study and pointed out that ICT infrastructure had different influences on its application for schools in urban and rural areas based on stepwise regression analysis (Chun, Chin-Chung, & Di, 2015).

More previous studies have mainly focused on the influencing factors of ICT application. However, relatively fewer empirical studies have been conducted to explore the relationship between the ICT application and ICT policy (about infrastructure, teachers’ training and technical support of ICT) in developing areas or even depressed areas on a nationwide scale. Therefore, this paper intended to study ICT application and to explore the correlation between ICT application and policy so as to identify the types of ICT policy to be constructed in cities and rural areas in national or provincial plans in education. We divided these variables into two groups: one is about ICT application: some variables related in the process of the classroom integration. The other is about ICT policy, based on infrastructure, teachers’ training and technical support of ICT. Make the following assumptions:

- ICT application needs to be fully evaluated.
- ICT application is affected by ICT policy to a certain extent at the school-level.
- Influencing factors on ICT application for schools in urban and rural areas is different.

In this paper, we chose more different measurable indicators about ICT application, then developed and tested linear regression models for some factors that explained ICT application. Specially, try to explore the relationship between application and policy of ICT at the school-level. Firstly, an exploratory factor analysis was used to divide the indicators of ICT application into three dimensions: ICT application of digital educational resources, ICT application of online learning and ICT application of educational environment, then explored the relationship between application and policy of ICT in middle and primary schools, finally made a comparative analysis for influencing factors of ICT application between the middle and primary schools in urban and rural areas.

2. Method

2.1 Sample

In this large-scale survey, the Departments of Education in each city were asked to nominate a provincial coordinator. All middle and primary schools in the sample districts were asked to complete the questionnaire. The leader and head teachers of each school completed the questionnaire according to the current situation of ICT at the whole-school level and sent it back to the provincial coordinator. Finally, the coordinators in each city sent back all the questionnaires to the research group.

There were about 10 thousands middle and primary schools in the Guangxi province of China. Samples gathered for this research were 8411 questionnaires of middle and primary schools over 14 cities. In this paper, 8380 questionnaires were used. These samples were selected randomly. In the questionnaire, we provided an individual choice question to know whether the school was in city or in rural areas. Among all the valid samples, 1642 were schools in cities, accounting for 19.59% of the total, and 6738 schools were in rural areas, making up 80.41% of the total. This study selected nine indicators of ICT application, as well as 6 indicators related to school ICT policy, as shown in Table 1.

Teachers’ perceptions about what teaching and learning processes can be improved through the use of ICT (Sangrà, 2011). The teacher’s role and the teaching processes are worth to be noticed. We used “Means of ICT teaching processes”, “Type of digital resources” and “Number of ICT-aided courses” to indirectly examine teachers’ application status of ICT. “Proportion of ICT-aided teachers” could directly reflect teachers’ application ability of ICT. The use of a web-based “teaching-learning space” to facilitate asynchronous interaction between students and their supervisors (Jowallah, 2014). Teachers and students can use online learning space to achieve teaching and learning. Thus, “Ratio of student online learning space” and “Ratio of teacher online learning space” were considered. Multimedia classroom represents one of the basic ways of conventional e-learning implementation (Zhao & Jiang, 2010). The main places where teachers integrate IT into classroom teaching are the
computer classroom, multimedia classroom (Yeh, Chang & Chang, 2011). Therefore, “Utilization rate of multi-media classrooms” and “Utilization rate of computer classrooms” were considered in our indicators. Moreover, according to the Compendium of Chinese ICT course in Middle and Primary Schools, the ICT courses in those schools comprise no fewer than 68 class hours per academic year, and the courses performed on computer are no less than 70% of the total lessons (MOE, 2010). Thus, “Time of ICT used in school per student” in China could reflect the status of students’ ICT utilization by multimedia classrooms and computer classrooms at the school level.

Table 1: The indicators of ICT application and policy.

<table>
<thead>
<tr>
<th>ICT application</th>
<th>ICT policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Means of ICT teaching processes</td>
<td>1. Proportion of Multi-media classrooms (P.M)</td>
</tr>
<tr>
<td>2. Type of digital resources</td>
<td>2. Number of Terminals per hundred students (N.T)</td>
</tr>
<tr>
<td>3. Number of ICT-aided courses</td>
<td>3. Proportion of Trained teachers (P.T)</td>
</tr>
<tr>
<td>5. Ratio of student online learning space</td>
<td>5. Proportion of ICT Supported teachers (P.S)</td>
</tr>
<tr>
<td>6. Ratio of teacher online learning space</td>
<td>6. Ratio of ICT Cost (R.C)</td>
</tr>
<tr>
<td>7. Utilization rate of multi-media classrooms</td>
<td></td>
</tr>
<tr>
<td>8. Utilization rate of computer classrooms</td>
<td></td>
</tr>
<tr>
<td>9. Time of ICT used in school per student</td>
<td></td>
</tr>
</tbody>
</table>

2.2 Data Analysis

Data were analyzed inferentially using SPSS17.0. The inferential statistics included an exploratory factor analysis, a Pearson’s correlation analysis and a series of regression analyses. Firstly, in order to reduce the data dimension of ICT application, exploratory factor analysis was conducted. Then a Pearson’s correlation analysis was carried out to examine the relationships among the variables of interest. Finally, the ICT policy factors were considered as predictor variables, whereas the three dimensions of ICT application were processed as outcome variables, and the different important factors that influenced ICT application in education in schools of different regions were analyzed according to a series of linear regression analysis results.

3. Results and Discussion

3.1 Factor Analysis

Considering the possible existing correlation between all indicators of ICT application, as well as the purpose of data dimensionality reduction from the large amount of data used, an exploratory factor analysis was used.
3.1.1 The Appropriateness Test

To determine the appropriateness of data and measure the homogeneity of variables entered into the analysis, the Kaiser–Meyer–Olkin (KMO) and Bartlett’s Test of Sphericity (BTS) were applied (Table 2). The KMO was 0.803, indicating that the sample was adequate for factor analysis (Kaiser, 1974). The BTS was 20100 ($p < 0.01$), indicating that the hypothesis variance and covariance matrix of variables as an identity matrix were rejected; therefore, the data were appropriate for factor analysis.

Table 2: Kaiser–Meyer–Olkin and Bartlett’s Test.

<table>
<thead>
<tr>
<th>Kaiser–Meyer–Olkin Measure of Sampling Adequacy</th>
<th>Bartlett’s Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.803</td>
<td>Approx. Chi-Square 20100 df 36 Sig .000</td>
</tr>
</tbody>
</table>

3.1.2 Factors Extraction

The Kaiser criterion was utilized to arrive at a specific number of factors to extract. Based on this criterion, only factors with eigenvalues greater than 1 were retained. Accordingly, three factors with eigenvalues over 1 were extracted, explaining a total of 63.84% of the variance. As shown in Table 3.

Table 3: Total Variance explained.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial Eigenvalues</th>
<th>Rotation Sums of Square Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>%of Variance</td>
</tr>
<tr>
<td>1</td>
<td>3.35</td>
<td>37.25</td>
</tr>
<tr>
<td>2</td>
<td>1.39</td>
<td>15.44</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>11.16</td>
</tr>
</tbody>
</table>

3.1.3 Factors Rotation

The varimax rotated factor analysis is shown in Table 4. The factors were named according to the contents of the relevant indicators. In the factor analysis, the percentage of the variance explained by each factor indicates the relative significance of the factor. The first factor explained 27.83% of the total variance, having a greater significance than the other two factors. This factor contains “Means of ICT teaching processes”, “Type of digital resources”, “Number of ICT-aided courses” and “Proportion of ICT-aided teachers”. These four indicators are based on educational resources to integrate ICT into education. Thus, a relevant name for the first factor on the loading pattern is “ICT application of digital educational resources”.

The second factor contains “Ratio of student online learning space” and “Ratio of teacher online learning space”. They are open proportion of student/teacher online learning space, which reflect the basic application of online learning for teachers and students to a certain extent. Therefore, according to these indicators a logical name that can be assigned to the factor is “ICT application of online learning”. This factor explained 18.88% of the total variance.

The last factor was associated mostly with variables related to “Utilization rate of multi-media classrooms”, “Time of ICT used in school per student” and “Utilization rate of computer classrooms”. The factor explains 17.13% of the total variance. “Time of ICT used in school per student” is the class time of students applied ICT every week, generally, in China multi-media classrooms and computer classrooms are places using ICT for students, so this factor can be named “ICT application of educational environment”.

According to the proportion accounted for the ICT application in middle and primary schools, we can infer that the three factors are ranked as follows: ICT application of digital educational resources, ICT application of online learning and ICT application of educational environment, among which ICT application of digital educational resources is the largest, and the other two are relatively smaller.
Table 4: Indicators loaded in the factors of ICT application using varimax rotated factor analysis.

<table>
<thead>
<tr>
<th>Name of factor</th>
<th>Item</th>
<th>Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Factor 1</td>
</tr>
<tr>
<td>digital educational resources</td>
<td>Means of ICT teaching processes</td>
<td>.856</td>
</tr>
<tr>
<td></td>
<td>Type of digital resources</td>
<td>.834</td>
</tr>
<tr>
<td></td>
<td>Number of ICT-aided courses</td>
<td>.785</td>
</tr>
<tr>
<td></td>
<td>Proportion of ICT-aided teachers</td>
<td>.588</td>
</tr>
<tr>
<td>online learning</td>
<td>Ratio of student online learning space</td>
<td>.825</td>
</tr>
<tr>
<td></td>
<td>Ratio of teacher online learning space</td>
<td>.812</td>
</tr>
<tr>
<td>educational environment</td>
<td>Utilization rate of multi-media classrooms</td>
<td>.767</td>
</tr>
<tr>
<td></td>
<td>Time of ICT used in school per student</td>
<td>.748</td>
</tr>
<tr>
<td></td>
<td>Utilization rate of computer classrooms</td>
<td>.570</td>
</tr>
</tbody>
</table>

3.2 Regression Analysis

3.2.1 ICT application in Middle and Primary schools

Based on the statistical analysis of all samples, the correlation between the schools’ application of ICT and policy of ICT was presented in Table 5. For middle and primary schools, it was found that all indicators about ICT policy were significantly related to selected ICT application, such as “Proportion of Multi-media classrooms” and “Proportion of ICT Full-time teachers” (r = 0.45, p < 0.01, r = 0.39, p < 0.01, respectively).

Table 5: Correlation between ICT application and ICT policy in middle and primary schools.

<table>
<thead>
<tr>
<th></th>
<th>P.M</th>
<th>P.F</th>
<th>P.T</th>
<th>P.S</th>
<th>R.C</th>
<th>N.T</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT application in middle and primary schools</td>
<td>.450*</td>
<td>.390*</td>
<td>.370*</td>
<td>.286*</td>
<td>.233*</td>
<td>.117*</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.01 level (2 tailed).

As also shown in Table 6, the indicators that were highly predictive for ICT application included the following: “Proportion of Multi-media classrooms” (t=24.68, p<0.001), “Proportion of Trained teachers” (t=18.35, p<0.001), of which the overall prediction proportion reached 31.40%. And also all indicators about ICT policy were significantly in predicting ICT application in middle and primary schools. The results showed that ICT policy could influence ICT application in middle and primary schools to the extent.

In order to explore how the ICT policy influenced ICT application in middle and primary schools, this study established regression models based on the all indicators about ICT policy, and a series of comparative analyses in urban and rural areas were conducted.

Table 6: Regression between ICT application and ICT policy in middle and primary schools.

<table>
<thead>
<tr>
<th>Dependent</th>
<th>Independent</th>
<th>B(st.d.)</th>
<th>S.E</th>
<th>T</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT application in middle and primary schools</td>
<td>P.M</td>
<td>.303</td>
<td>.007</td>
<td>24.68*</td>
<td>.314</td>
</tr>
<tr>
<td></td>
<td>P.F</td>
<td>.167</td>
<td>.007</td>
<td>13.16*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.T</td>
<td>.212</td>
<td>.006</td>
<td>18.35*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.S</td>
<td>.122</td>
<td>.006</td>
<td>10.25*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R.C</td>
<td>.078</td>
<td>.006</td>
<td>6.90*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N.T</td>
<td>.033</td>
<td>.005</td>
<td>2.96*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Constant)</td>
<td>.099</td>
<td>.006</td>
<td>15.94*</td>
<td></td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.01 level (2 tailed).
3.2.2 Analysis on the Influencing Factors of ICT Application’s Dimensions of Urban and Rural Areas

For middle and primary schools in urban areas, it was found that “Proportion of Trained teachers” was significantly related to both online learning and educational environment (r = 0.312, p < 0.01, r = 0.254, p < 0.01, respectively) of ICT application. As for rural schools, ICT application with the highest degree of correlation to the ICT policy was “Proportion of Multi-media classrooms” (r = 0.376, p < 0.01). As shown in Table 7.

In short, there was a certain degree of correlation between ICT application and ICT policy in schools of all regions. We found that there was a correlation between “Proportion of Multi-media classrooms” and “ICT application of digital educational resources” no matter schools in urban area or rural area, and the former was less significant than the latter. However, compared with urban area, a correlation not only existed between “Proportion of Multi-media classrooms” and “ICT application of digital educational resources” in rural area but also significant.

Table 7: Correlation between ICT application and policy in the two regions.

<table>
<thead>
<tr>
<th></th>
<th>Dependent</th>
<th>P.M</th>
<th>N.T</th>
<th>P.T</th>
<th>P.F</th>
<th>P.S</th>
<th>R.C</th>
</tr>
</thead>
<tbody>
<tr>
<td>urban area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT application of digital educational resources</td>
<td>.187**</td>
<td>.062*</td>
<td>.254**</td>
<td>.213**</td>
<td>.154**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT application of online learning</td>
<td>.105**</td>
<td>.088*</td>
<td>.312**</td>
<td>.080**</td>
<td>.119**</td>
<td>.186**</td>
<td></td>
</tr>
<tr>
<td>ICT application of educational environment</td>
<td>.196**</td>
<td>.054*</td>
<td>.080**</td>
<td>.170**</td>
<td>.103**</td>
<td>.010**</td>
<td></td>
</tr>
<tr>
<td>rural area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT application of digital educational resources</td>
<td>.376**</td>
<td>.074*</td>
<td>.225**</td>
<td>.179**</td>
<td>.139**</td>
<td>.222**</td>
<td></td>
</tr>
<tr>
<td>ICT application of online learning</td>
<td>.257**</td>
<td>.055*</td>
<td>.065**</td>
<td>.190**</td>
<td>.083**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT application of educational environment</td>
<td>.145**</td>
<td>.089*</td>
<td>.223**</td>
<td>.100**</td>
<td>.131**</td>
<td>.164**</td>
<td></td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2 tailed). *.Correlation is significant at the 0.05 level (2 tailed).

As shown in Table 8, according to the results of the regression analysis, for schools in rural area, 20% of ICT application of digital educational resources were explained by ICT policy to the model significantly (R² = .200), and 8.8% of ICT application of educational environment were explained by ICT policy to the model significantly (R² = .088), which were both higher than those in urban area (R² = .116, R² = .059). However, for schools in urban area, 13.2% of ICT application of online learning was explained by ICT policy to the model significantly (R² = .132), which was higher than that in rural area (R² = .082). In fact, rural middle and primary schools faced a lot of difficulties. In the term of hardware of ICT, there were a small number of computers and multimedia facilities with low configuration. In the aspect of software use, no matter the quantity or quality, educational resources were difficult to meet the normal needs. In the aspect of teacher staff construction, many problems existed, such as: less ICT full-time teachers, the heavy workload, the low level of ICT application and the lack of opportunities participating in ICT training. In order to further find out the different influencing factors of ICT application in two regions, more comparative analyses were completed from the following three application dimensions:

In ICT application of digital educational resources, for schools in cities, the indicator “Proportion of Trained teachers” (t=8.70, p<0.001) related to school ICT policy was highly predictive. Urban schools could use rich digital educational resources, but resources updated faster, which gave teachers a big challenge about how to integrate ICT into classroom well. ICT training was an effective way to promote teacher professional development, and it could improve teachers’ ICT ability (Sang, Valcke, van Braak, Tondeur, & Zhu, 2011). According to participate in ICT-related training, teachers could deepen the understanding, and applied it to the actual teaching process, so as to better promote the teaching and learning. For schools in rural areas, the indicators related to school ICT policy that were highly predictive for ICT application of digital educational resources included the following: “Proportion of Multi-media classrooms” (t=26.59, p<0.001) and “Ratio of ICT Cost” (t= 13.62, p<0.001). The hardware facilities of rural schools were not popular, research data showed that about 34.23% of the schools failed to build multimedia classrooms. Different from the schools in cities - the
use of mobile terminals and the future classroom had become increasingly popular, multimedia classroom was still the main place of digital teaching in rural schools. The hardware equipment and network building in rural schools have been relatively inferior, restricting the further application of digital educational resources (Jingtao, Yuanjuan, & Xiaoling, 2010).

In ICT application of online learning, for schools in cities, the indicator——“Proportion of Trained teachers” (t=11.78, p<0.001) related to ICT policy was highly predictive. With the innovation of classroom teaching, the teaching mode in the urban schools was gradually transformed from “teacher-centered” to “student-centered”, requiring teachers to be provided with teaching skills, especially taking advantage of online learning space to promote teaching and learning. For schools in rural areas, the indicators related to school ICT policy that were highly predictive for ICT application of online learning include the following: “Proportion of Multi-media classrooms” (t=17.26, p<0.001) and “Proportion of ICT full-time teachers” (t=9.65, p<0.001). The limited number of multimedia classrooms, the lack of advanced terminals and the lack of targeted professional teachers were the main problems of ICT online learning in rural schools.

In ICT application of educational environment, for schools in cities, the indicators related to school ICT policy that were highly predictive include the following: “Proportion of Multi-media classrooms” (t=6.41, p<0.001) and “Proportion of ICT full-time teachers” (t=4.29, p<0.001). With the continuous education reform in urban school, the application of multimedia equipment has become the normalization of classroom, which put forward a high standard to the ICT use of teachers. With the rapid development of ICT in education, the ICT use of teachers has become more specialized, ICT full-time teachers could depend on professional knowledge about multi-medium, computers and other equipment. For schools in rural areas, the indicators related to school ICT policy that were highly predictive for ICT application of educational environment included the following: “Proportion of Trained teachers” (t=14.63, p<0.001) and “Ratio of ICT Cost” (t=9.01, p<0.001). Rural schools mainly faced with a lot of difficult issues, such as fewer ICT professional staff, the lack of teachers who owned awareness and less cost of ICT-related training.

Generally speaking, ICT application in urban and rural middle and primary schools were at different stages, and there were differences in ICT application respectively. Under the premise of the improvement of infrastructure, to keep up with the pace of ICT in education, urban areas were experiencing the change—from “how to use” to “how to make good use”, which put forward a higher level of professional requirements for ICT towards teachers. In comparison with city schools, ICT in rural schools was starting from scratch, ICT application still depended on the complete degree of facilities. Lacking of ICT cost in rural areas, the investment in the introduction of advanced equipment and the construction of teaching staff were less than the city.

Table 8: Regression between ICT application and policy for middle and primary schools in two regions.

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Predicting variables</th>
<th>Urban areas</th>
<th>Rural areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT application of digital educational resources</td>
<td>P.M</td>
<td>.115</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>P.T</td>
<td>.209</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>P.F</td>
<td>.121</td>
<td>.014</td>
</tr>
<tr>
<td></td>
<td>P.S</td>
<td>.095</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>R.C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT application of online learning</td>
<td>P.M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N.T</td>
<td>.052</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>P.T</td>
<td>.281</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>P.F</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.S</td>
<td>.138</td>
<td>.023</td>
</tr>
<tr>
<td>ICT application of educational environment</td>
<td>P.M</td>
<td>.160</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>N.T</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.T</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.F</td>
<td>.111</td>
<td>.026</td>
</tr>
<tr>
<td></td>
<td>P.S</td>
<td>.051</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>R.C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Conclusion

The present study focused on the influencing factors of ICT application in middle and primary schools. It was important to stress that the study was set up in the Chinese context. Our findings suggested that the influencing effect of school ICT policy to ICT application was different in urban and rural areas. The results underpin the importance to consider linear regression models to explain the influencing factors and predict ICT application.

Using a large representative dataset of 8380 schools in the Guangxi province of China, a series of regression prediction models were constructed to analyze the influence that ICT policy may exert on ICT application. Based on the prediction results of the models and the comparison in two religions, suggestions for the policy of ICT in the two regions were provided for further popularizing ICT application in schools. In order to speed up the process of ICT application in schools, taking the gap in the development and application of ICT between two religions into account, combined with the different stages of ICT application respectively, our suggestions for ICT policy in different regions were as follows:

For urban schools, it’s necessary to strengthen the construction of teachers and enhance the information literacy of teachers (Álvarez & Gisbert-Cervera, 2015). The construction of teachers was the basic guarantee for the sustainable ICT development in education, and the ICT application ability was the necessary professional ability for teachers in modern society. Teachers’ ICT status and ICT training were the relevant contents of the teacher staff construction, with the continuous emergence of high-quality digital educational resources, schools in cities should strengthen teacher training (Tsiouridou & Vryzas, 2004) related ICT, and constantly improve the ICT application level of teachers to help them integrate ICT into classroom perfectly, achieving a wide range of digital teaching methods, so as to promote teaching and learning.

For rural schools, due to lack of ICT cost in middle and primary schools, ICT infrastructure was seriously lagging behind developed areas such as schools in cities. Therefore, rural schools needed to increase the following three aspects of ICT investment:

- In the development of ICT in middle and primary schools, infrastructure construction was the prerequisite for ICT application, so it’s necessary to improve infrastructure construction, such as build enough multimedia classrooms and provide better ICT equipment, so as to satisfy the normal teaching needs.
- Considering ICT in education centering on its application, to realize ICT application needed to have a better hardware environment as a support. More importantly, there must be rich to meet the needs of high-quality digital education resources. To achieve the purpose of resource sharing, more high-quality digital educational resources should be purchased, so as to bridge the “digital divide” with urban areas.
- Finally, organize the ICT training actively, encourage teachers to participate in it, provide teachers with the opportunities of ICT application and develop more ICT full-time staff, so as to provide technical guidance (Papanastasiou, 2008) for the future.

In recent years, ICT in education of China has been rapidly developed. Infrastructure construction was generally completed in urban areas, rural areas still remained at a fairly junior level. Shortage of ICT cost were usually in resource construction and teachers’ training. Most teachers paid attention to the latest development of technology and had a certain degree of computer application ability. However, there were many problems, such as the heavy task, the lack of a large number of outstanding professional teachers and lagging teaching mode. At present, ICT in education of China is in the initial stage. As a developing country of the Third World, although there are many difficulties in the process of ICT, China is still drawing on the ICT system put forward by developed countries, continues to vigorously promote ICT application combined with own actual situation as well. We hope the way China coped with can provide the developing countries even underdeveloped areas, such as the African region with some feasible ideas.
Acknowledgements

This work was supported by National Social Science Fund Program for Project: “Research on the Differences and Influencing Factors of Urban and Rural Information in Middle and primary schools” under grant No.16CGL058.

References


PIVOTeeING: A Flipped Approach in a Postgraduate Solid State Devices Course

Lakshmi T Ga*, Soumya NARAYANA a, Harshvardhan PENUGONDA a, Dhirendra VAIDYA b, Vishvendra POONIA a, Swaroop GANGULY b & Sahana MURTHY a

*IDP in Educational Technology, IIT Bombay, India
bDepartment of Electrical Engineering, IIT Bombay, India
*tglakshmi@iitb.ac.in

Abstract: Design and implementation of flipped courses is a challenge as it involves many elements. Flipped approaches, though well-suited for engineering courses, have not been implemented and reported as much as K-12. A review of the various meta-studies also points out that due to the presence of complex activities in flipped approach instructors need support for the design, implementation and evaluation of flipped classrooms. In this paper, we describe the design, implementation and evaluation of a flipped approach for a postgraduate Electrical Engineering course - Solid State Devices. The instructors and educational researchers from our institute worked in a partnership framework for the flipped approach - Problem solving Via instructOr mediaTed peer learING (PIVOTeeING). Since it was an exploratory cycle in the Design Based Implementation Research (DBIR) framework, various learner affect and performance parameters were explored. The results show that, for more than 65% of the time, learners were engaged with the content and more than 60% of learners attained satisfactory levels in an end-semester exam on a new concept not specifically addressed in class. The implementation was also carefully examined to present practice-based guidelines for implementation of flipped classroom in the electrical engineering domain via the participatory design framework.

Keywords: flipped classroom, participatory design, solid state devices

1. Introduction

A common strategy being implemented in current courses is the flipped classroom, where learners are introduced to course content outside of classroom and provided with deeper engagement opportunities within the class. The flipped classroom design is equivalent to the design of a system consisting of different moving parts (Talbert, 2017). The core elements of this flipped design are: (i) content exposure prior to class, (ii) mechanism to assess out-of-class learning, (iii) higher-level cognitive activities in-class (Brame, 2013). The design rationale behind converting any course to flipped format is to utilize the class-time for engaging learners in higher order cognitive activities (Velegol, Zappe & Mahoney, 2015).

The implementation of the flipped approach has predominantly been reported in K-12 education, for example, a meta-review on flipped classroom revealed that 90% studies (of about 759 original studies reported from 1995 to 2015) were in the K-12 setting (O’Flaherty & Philips, 2015). However, this pedagogical approach is gaining popularity in higher education as well. 164 flipped studies in engineering were reported from 2000-2015. (Karabulut-Ilgu et al., 2017). Flipped learning appears to be well suited for engineering education as it “combines learning theories once thought to be incompatible—active, problem-based learning activities founded upon a constructivist ideology and instructional lectures derived from direct instruction methods founded upon behaviorist principles” (Bishop & Verleger, 2013). In this paper we describe the course design, implementation and evaluation of a flipped classroom PIVOTeeING (Problem solving Via instructOr mediaTed peer learING), for the course ‘Solid State Device’ for the post-graduate program in Electrical Engineering at IIT Bombay.

The following recurring recommendations emerge from various meta-reviews of the flipped classroom: (Karabulut-Ilgu et al., 2017; O’Flaherty and Philips, 2015; Kim et al, 2014; Howitt, Christine and Pegrum, 2015; Lo et al., 2017)
i. Flipped approach needs more careful implementation and reporting to understand how and which of its elements contribute to learning

ii. Flipped course practitioners need support to create learning sequences, use of technology in learning intervention

iii. Flipped courses need to be evaluated for student learning effectiveness for which the practitioners need help in design, implementation and evaluation.

Instructors need support for implementing the above recommendations. At the same time, research is required to explain the working of a flipped approach in a way that informs the practice. One strategy that has been recommended to address these goals is a collaboration between educational researchers and practitioners (Peneul et al, 2015). The researcher’s role is to translate their work on learning to interventions implementable by instructors and conduct educational evaluation, while the practitioner implements the learning interventions in continuously evolving environments. This collaboration will ensure that the research builds closely from practice and would be relevant to the implementation. The practitioners would deepen and enhance their own practice by engaging in the research process. A research-practice partnership towards the design, implementation and evaluation of flipped approach could systematically provide answers to the questions arising from the meta-reviews.

Thus, the other goal of this paper is to discuss the interactions between the researcher team and the practitioners (which included a faculty instructor and 3 TAs). Based on our participative evaluation design, we present the outcomes of this flipped implementation in the form of observations of in-class engagement, learners’ perception and learners’ end semester scores. We have tried to reflect post-facto how we worked and what all can be improved in-terms of the flipped approach design as well as the collaboration. Based on the collaborative effort we have come up with guidelines at different levels and in terms of different aspects to be considered while taking a flipped approach to a course. As with any educational strategy, reports on evidence based flipped classroom practices will contribute towards a theoretical framework that may be able to answer the ‘why’ and ‘how’ questions about the strategy. Such research-based practices would benefit the primary stakeholders in the classroom - instructor as well as learners.

2. Related Work

2.1 Flipped Classroom

A flipped classroom has been defined as a combination of interactive group learning activities in the classroom, and computer-based instruction outside the class (Bishop & Verleger, 2013). This approach differs from the traditional approach, where most of the instruction is delivered in class and most of the learning activities are given to the students outside the class. While the traditional approach has historically been effective in disseminating large amounts of course content to a large audience, there is mounting evidence this approach does not necessarily promote a high level of learning or retention of knowledge (Borrego & Bernhard, 2011). Most of the in-class time is spent by the student listening to the instructor and taking notes. Instructors do spend time in higher order thinking activities like problem solving, discussions etc during class, but this can be done only after relevant content has been covered. Traditional lectures are not self-paced, and many students require more time to understand a particular concept. Flipping the classroom can mitigate the above difficulties to a large extent. Students engage with the basic concepts outside class at their own pace. Hence a lot of in-class time can be devoted to higher order thinking activities. Difficulties and misconceptions can also be addressed by the instructor in-class.

Studies reporting the flipped classroom approach are on the rise in the past few years (Giannakos et al, 2014). The flipped classroom approach has been used in school-level, undergraduate as well as graduate courses, covering a wide range of disciplines like Biology, Physics, Architecture, Computer Science, Electrical engineering, Mechanical Engineering, Clinical Nursing, Management, Pharmacy, English speaking, History, Mathematics and Sociology (Giannakos et al, 2014; O’Flaherty & Philips, 2015).

Most studies measure the learning gains and students’ perception of the flipped classroom approach. Findings of a meta-study (O’Flaherty & Philips, 2015) suggest that the flipped classroom approach shows increased academic performance compared with historical controls. However, very
few studies evaluated increase in students’ higher order thinking skills, such as problem solving, inquiry and critical thinking. There is also limited published evidence on long term student learning outcomes. In another meta-study (Giannakos et al, 2014), the authors report that more than half of the studies analyzed suggest that students have positive attitudes toward the flipped classroom approach. Students especially liked that they could learn independently and at their own pace. The authors suggest that the active learning component of flipped classroom enabled students to remain at higher order levels of the Bloom’s Taxonomy for longer durations of time, and hence students perceive that the quality of learning is better.

There are certain reported critiques of the flipped approach as well. The in-class component requires students be actively engaged and hence more accountable. This more active role is difficult for some students to adjust to (Bormann, 2014). In one of the studies (Mason, Shuman & Cook, 2013), based on student feedback, the authors conclude that this format might not work in a course with many new concepts because students would struggle to identify where to apply the various new concepts and equations. Also, students disliked video lectures and claimed that recorded lectures are not appropriate for more difficult course material (Giannakos et al, 2014). The flipped classroom approach may not be applicable to all subjects and also depends on the willingness of learners to adapt to the new format. The meta-study by O’Flaherty and Philips (2015) also states existing studies do not have robust evidence to show that the flipped classroom approach is more effective than conventional teaching methods.

2.2 Participative Design

Course instructors and educational researchers are members of different but related communities of practice (Goos, 2014). The members of the community working in isolation may not result in improvement of the practice. It becomes essential that they work together. While working together might seem as the solution, the relationship needs to be based on a partnership collaboration model. Participative design is the approach of design characterized by stakeholder involvement (Spinuzzi, 2005). Participative design draws on various research methods such as ethnographic observations, protocol analysis, interviews, and analysis of artifacts. These methods are used iteratively to construct the design, which is emerging. The design is then implemented on field after which the results are co-interpreted by the designer-researchers and the users of the design. Participative design is based on partnership with the stakeholders which needs to be iteratively conducted.

There exists different models of researcher-practitioner partnership, in which practitioners: (i) either participate in research as objects of inquiry or (ii) work alongside researchers engaging in form of inquiry. In this paper we were working on designing a flipped approach towards the teaching-learning of post graduate electrical engineering course Solid State Devices, hence we look at the researcher-practitioner partnership engaging in a form of inquiry. PIVOTeeING started with design by the practitioner after which for the implementation and evaluation the practitioner partnered with the researcher. The aim of the evaluation is to feed back into the design.

3. Design and Implementation of PIVOTeeING

PIVOTeeING – Problem solvIng Via instructOr mediaTed peer learING - was designed for the course – Solid State and Devices. This is a compulsory (elective) course for the Master’s (Ph.D.) program in the Microelectronics specialization within Electrical Engineering. It is expected that after going through the course student would be able to (i) judiciously choose devices for a design, (ii) verify device behavior experimentally, (iii) analyze different devices with different compositions, (iv) tweak existing device structure and composition to obtain a desirable behavior and (v) create a new device with certain specific desirable characteristics and behavior. At a broad level the design of PIVOTeeING is based on the general flipped classroom guidelines (Brame, 2013). For the in-class activities the instructor chose the active learning strategy of Peer-Instruction (Mazur, 1997).

The instructor had developed a set of learning objectives to achieve the course outcomes. The learners would need to solve solid state devices problem by reasoning: (i) mathematically, (ii) conceptually, (iii) diagrammatically and (iv) connect the three of them. These objectives were crucial to the design of flipped approach as the in-class activities were designed based on these objectives. Similar to many flipped models, PIVOTeeING design also had pre-class, in-class and post-class
elements. The pre-class element was asynchronous which would enable the learner to learn content in their own pace. During the class, students were involved in problem solving with peers, mediated by the instructor. At this phase they were also encouraged to view multiple representation (MR) such as mathematical equation and the corresponding band diagram. Fig. 1 explains the design in detail.

![Figure 1](image.png)

**Figure 1.** PIVOTeeling (Problem solving Via instructOr mediaTed peer leanING): A flipped approach

### 3.1 Course Materials

For the pre-class phase, the instructor had prepared screen casts with audio narration. The presentations were uploaded on Moodle learning management system. All courses in our institute are run on Moodle, so the learners had experience using the platform. At the end of presentation, the instructor posed a problem, which the learners were required to solve before they came to class. All pre-class materials were made available on Moodle as well. Any course related announcements, in terms of quizzes or assignment submissions were also made on Moodle.

For the in-class activities, the instructor implemented a series of active learning and collaborative strategies. The most common activity was Peer Instruction (Mazur, 1997) in which the instructor posed a conceptual multiple choice question and students had to vote and re-vote after peer discussion. The questions started from explain level in Bloom’s taxonomy and moved towards higher levels. An example of a question in Peer Instruction activity is:

“The potential profile across a device is shown in the above graph. The total space charge inside the device is: a) Positive, b) Zero, c) Negative, d) Cannot be determined from the given information”

At times, activities such as Think-Pair-Share were implemented, for questions that required deeper and longer reasoning. The instructor mediated the discussion and provided hints and additional information as required. All activities entailed learners to perform actions based on the learning objectives (Section 3). The instructor also posed an open-ended question at the end of class for the learners to explore post-class.

### 3.2 Interaction between Instructor and Researchers

The course instructor designed and implemented the flipped approach. The instructor had previously watched online lectures from the Educational Technology department in our institute, regarding active learning pedagogies. The videos had guidelines for practitioners to implement the active learning pedagogies such as Think-Pair-Share and Peer Instruction. Based on the guidelines the instructor along with the course Teaching Assistants prepared a multiple-choice question bank for the whole course. The options for each multiple choice question were carefully designed to explore all the solution paths and scenarios. The instructor created peer instruction activity for the learners based on the questions.

While the instructor implemented the flipped design in-class, the research team performed in-class observations to capture the learner engagement during the activities. The research team used a
previously validated observation protocol (Kothiyal et al, 2013) to observe in-class learner engagement. The observation protocol had a list of engaged on-task, engaged off-task and completely disengaged behaviors. A team of five researchers were present in-class for observations. Each of the researcher observed 10 randomly picked students throughout for 2 classes.

The instructor had also prepared formative quizzes to understand the performance of learners. The formative quizzes were administered for 30 minutes after class. While evaluating the quizzes, the research team initially found that learners were not in line with the expected solution. In the words of the instructor – “It appears that the quizzes were far too tough”. The detailed analysis of the quizzes revealed that they did not provide explicit mechanisms for learners to apply what they had practiced in class. So the research team suggested alignment of quiz questions to activities in class, especially for activities at the Analyze and Evaluate levels. The instructor adopted the suggestions during preparation of the end-semester exam, which was then evaluated by the instructor, TAs and the research team.

4. Research Methodology

PIVOTeeING was a collaboration of partnership between the two communities of practice (Electrical Engineering educators and Educational Technology researchers) existing within our institute. The teams partnered to systematically design, implement and evaluate the flipped approach. We worked in the framework of Design Based Implementation Research (DBIR). DBIR applies design-based perspectives and methods to address and study problems of implementation (Fishman et al, 2013). The core principles of DBIR are: (1) a focus on persistent problems of practice from multiple stakeholders’ perspectives; (2) a commitment to iterative, collaborative design; (3) a concern with developing theory and knowledge related to both classroom learning and implementation through systematic inquiry; and (4) a concern with developing capacity for sustaining change in systems. This paper reports the first cycle of the DBIR implementation.

4.1 Research Questions

We took an explorative approach to examine various aspects of the effect of the flipped classroom. Since learners are the primary stakeholders we first examined the impact on the learners such as affect and learning outcome. The RQs related to the learners are:

RQ1: How did the implementation of PIVOTeeING impact learners' affective parameters such as engagement and perception?

1.1. How engaged are the learners during in-class activities?
1.2. What is the learner perception about various design elements of the course?

RQ2: How did the learners perform in the end semester exam with respect to the learning objectives?

The other goal of this paper is to evaluate the design of PIVOTeeING and provide guidelines to the practitioners via a researcher-practitioner partnership. The RQ related to this objective are:

RQ3: What changes are required for the design, implementation and evaluation of PIVOTeeING flipped approach?

4.2 Participants

The class consisted of a total of 58 students. Among these students there were 2 cohorts, the postgraduate and doctoral students. Among the 58 students only 36 agreed to participate in the study. However, only 32 of the 36 students appeared for the end semester examination. All of the students were from the electrical engineering background with a basic exposure to solid state device course.

4.3 Instruments

4.3.1 Observation Protocol

The in-class activities of PIVOTeeING aimed at learners working at higher order levels in Bloom’s taxonomy. The assumption here is that by doing so the learners would be actively engaged with the
content of the course for most of the time in-class. To measure the in-class engagement, we used the real-time observation protocol for learner engagement behaviours by Kothiyal et al (2013). Since the instrument was already validated in previous studies, we checked the reliability of the observations. The research team had five observers. All of the observers picked a set of same 10 students in the class and performed observations throughout the class. The team observed each student for 10 seconds. In the first round of observations, there was only 36% agreeability, as there was no synchronization in the observation timing. For the next round, the team synchronized the observation timing by reducing the observation timing to 5 seconds and recording only the dominant behavior. After the second round of observations there was agreement of 81%.

4.3.2 Rubrics for End-semester Exam

The end semester exam had 3 questions which were evaluated at two levels. The first round of evaluation was done by the TAs. The second round of solution analysis was done by the research team. This evaluation was more fine-grained and the intention was to look at how much learners were able to achieve the learning objectives (Section 3). A six criteria rubric with 3 scales was developed by the research team. These criteria were: C1 – Application of appropriate concepts of semiconductor devices, C2 – Identification of mathematical construct to solve the problem, C3 – Appropriate linkage between semiconductor concepts and mathematical constructs, C4 & C5 – construction of relevant diagrams, C6 – Linkage between concept, mathematical concept and diagram to devise solution.

The rubric was checked for content validity by checking alignment with the learning objectives and in-class activities. To ensure reliability of scoring using the rubric, two iterations of independent evaluation was done for 6 student answers. Cohen’s Kappa was calculated for each criteria in the rubric: \( \kappa_1 = 0.66, \kappa_2 = 1.0, \kappa_3 = 1.0, \kappa_4 = 1.0, \kappa_5 = 1.0, \kappa_6 = 0.64. \)

4.4 Data Analysis

The data analysis involved use of mixed methods. To answer RQ 1.1, quantitative methods such as aggregating the behaviors and scores was used. RQs 1.2 and 3 were answered using thematic analysis of two focus interviews. Focus group interviews were conducted for two cohorts of learners - one being post graduate students and the other being PhD students. The interviews were transcribed and two researchers independently coded the transcripts to elicit themes. The transcripts were analyzed using the guidelines provided by Clarke and Braun (Clarke and Braun, 2014). After the initial coding, the researchers compared their codes and categories, thereby refining them considerably in the process. The discussion also helped in reviewing the emergent themes. Fig. 5 (Section 5.2) summarizes the main themes, sub-themes and categories that emerged from the focus group interview data. Table 1 summarizes for each RQ (Section 4.2), the data sources, instruments and data analysis methods used.

Table 1: Research Questions - Method – Analysis

<table>
<thead>
<tr>
<th>RQs</th>
<th>Data Source</th>
<th>Instrument</th>
<th>Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1.1</td>
<td>In-class student behaviors</td>
<td>Observation protocol</td>
<td>Analyze engaged behaviors, aggregate the behaviors and compare engaged and non-engaged behaviors</td>
</tr>
<tr>
<td>RQ1.2</td>
<td>Transcript of focus group interview (FGI) with 2 cohorts of learners</td>
<td>-</td>
<td>Thematic analysis of the FGI transcript</td>
</tr>
<tr>
<td>RQ2</td>
<td>End semester exam answers</td>
<td>End semester evaluation rubric</td>
<td>Rubric based evaluation</td>
</tr>
<tr>
<td>RQ3</td>
<td>End semester exam answers and transcript of FGI</td>
<td>Rubric based evaluation and thematic analysis of the FGI transcript</td>
<td></td>
</tr>
</tbody>
</table>
5.1 Student in-class Engagement

The student engagement overall in class is presented in Fig. 2. From the engagement behaviour distribution, we see that apart from 15% of reading instructor presentation, the students are actively engaged with the content. Fig. 3 presents the engagement across the activities in one class. From the bar graph we see that for more than 65% of the class time (90 mins) learners were engaged with the content.

![Figure 2. Overall In-class Engagement](Image)

![Figure 3. In-class engagement across activities](Image)

5.2 Student Perception of Course Material and Learning Strategy

Analysis of the two focus group interviews reveal 2 major themes with 6 key sub-themes. The student related themes covers sub themes such as student needs, challenges and how they learn (Fig. 4). The other major theme was related to the course which covers sub-themes of perception about course content, benefits of course format and problems in course format (Fig. 5).

5.3 Achievement of Learning Objectives

Using the rubric, the end semester answers were evaluated (N=26). The end semester exam had three questions based on different devices Junction field-effect transistor -JFET (Q1), N metal-oxide-semiconductor field-effect transistor - NMOSFET (Q2) and PNP Bipolar Transistor (Q3). Q1 and Q2 had scaffolds (references to in-class activities) to answer the question whereas Q3 did not have any scaffolds. The rubric criteria applicable to all questions were C1, C2, C3 and C6. C4 and C5 were only applicable to Q2. Fig. 6 compares applicable criteria (C1, C2, C3, and C6) across the three questions. It is seen that across the criteria C1 and C2 the learners have performed well in Q1 (device JFET). It is to be pointed out that JFET as a device was not separately discussed in class. However, when students had to answer questions related to this device in the end semester exam, they had to synthesize from multiple concepts covered in class. In the answers for Q2 with regards to the criteria C4 and C5, only 15% (4) have performed well in both.

![Figure 4. Student related themes from FGI](Image)
6. Discussion and Conclusion

Research questions pertaining to engagement (RQ1.1) are answered by the pie-chart depicting ‘overall student engagement’ and bar graph of ‘engagement behaviors across activities’ (Figs. 3 and 4). The students were engaged with the content more than 65% of the class time. From the two charts we see that across the activities, students were mostly engaged in solving problem/questions individually and listening. Since most of the dominant behaviors were of writing and listening, in-class questions could include more scaffolds for peer discussion for focused discussion.

With regards to learner performance (RQ2), more than 60% of learners were able to achieve target level of concept identification and satisfactory level of mathematical construct recall for solution in the context of JFET and NMOSFET than PNP Bipolar transistor. 50% of learners are able to achieve satisfactory level of establishing link of mathematical construct and concept for JFET. This result may indicate transfer of learning, as JFET as a concept was not explicitly handled in class. Additionally the question with JFET and NMOSFET had scaffolds (reference to in-class activities) to answer the question whereas the PNP Bipolar transistor question did not have scaffolds. This seems to indicate that if the questions had sufficient scaffolds learners are able to apply their learning to an unknown concept. So PIVOTeeING has been able to achieve near transfer.

The design of PIVOTeeING included elements of course format, course material and evaluation. The themes related to student perception (RQ1.2) that emerged in the analysis of focus group interviews corresponded to these elements. Students perceived that while the problem solving approach and Peer Instruction activity were useful in exposing them to concepts with multiple viewpoints, they would like to have a largely lecture driven format. The reasons for this preference seem to stem from the deeply embedded bias for the traditional teaching format with a visible reluctance to embrace a new pedagogical format. Lack of opportunities for doubt clarification and mismatch between out-of-class and in-class activity seems to have added to student discomfort with the format. The unavailability of opportunities to bridge out-of-class and in-class content led to the student belief that short videos lacked depth. Students also felt that there was a disproportionately lesser allocation of marks for class quiz, homework and discussions, considering the effort that went towards them.

The themes of student needs, problems in course format, course content, how student learn, and student challenges, bring out the discomfort of the students about the different aspects of the PIVOTeeING method. Based on this, we have come up with guidelines (Table 2) to be applied for design and implementation of PIVOTeeING flipped approach. These guidelines will be carefully included in the design of PIVOTeeING for the second cycle of DBIR.
This study contains certain limitations. First, the RQ related to learning performance (RQ2) was a post-test only design. This design is known to be weak, and is at best indicative of the potential of the intervention. We were able to obtain such indications from the current study, however, a stronger research design is required to make claims of student performance from the PIVOTeeING flipped approach. Second, the observation protocol that was used to measure in-class student engagement (RQ1.1) needs to be refined. Since we adapted the instrument used in a different context (Kothiyal et al, 2014), it needs to be contextualized to certain activities specific in PIVOTeeING. For example, during in-class activities such as ‘reading’, we were not able to decode whether there is active engagement or not. Third, the implementation of the observation protocol was time and labor intensive, hence it limits scaling, especially in large classes. Since this first implementation was primarily an exploratory approach, we plan to use the insights from this cycle to systematically carry out an evaluation study in the next cycle. Third, there was a lack of measurement of learners’ performance in pre-class. We gathered insight about the learners’ video content viewing and learning time through a self-reported survey, but it does not directly indicate the pre-class learning performance.

Table 2: Guidelines for implementing PIVOTeeING flipped approach

<table>
<thead>
<tr>
<th>Design Element</th>
<th>Guidelines for flipped approach</th>
</tr>
</thead>
</table>
| Pre-Class Activities | • Create videos carefully with appropriate reflection spots, learning by doing activities and self-assessment quizzes  
• Incorporate formative feedback to learners during the pre-class activities  
• Provide platform to let learners discuss the doubts/queries in the video  
• Moderate the discussion platform for learner’s problems/misconceptions                                      |
| Bridge Activities    | • Create bridge activities in pre-class for learners to link to the in class activities  
• Summarize pre-class activities or conduct Q/A for pre-class activities  
• Create a focus question or problem on the class activities for post-class discussion  
• Provide platform to let learners discuss the focus question/problem post-class  
• Moderate the discussion for learner’s problems/misconceptions                                                                                      |

A contribution of this study is the implementation and evaluation of a flipped approach in post graduate engineering course using a participatory design approach. For practitioners who wish to implement a flipped approach, in addition to recommendations in Table 2, there are additional important recommendations related to affect: (i) get student buy-in, particularly for the non-traditional aspects of learning within a flipped class, (ii) seek training on aspects of implementation that you may not be familiar with— both technological (for ex creation of videos) and pedagogical (for ex conducting
PI) and (iii) be prepared to do two or more implementations before you start seeing results. Table 2 along with the affect related guidelines provide a practice based guideline for implementation of flipped classroom in electrical engineering domain via the participatory design framework. Finally further cycles’ implementation and evaluation would lead to the development of a framework to guide a practitioner to systematically flip any course.

References


Computational Thinking Development through Programmable Robotics Activities in STEM Education in Primary Schools

Siu-Cheung KONG * & Chan-Chio LAO
Centre for Learning, Teaching and Technology,
The Education University of Hong Kong, Hong Kong
*sckong@eduhk.hk; acclao@eduhk.hk

Abstract: Programmable robotics activities in science, technology, engineering and mathematics (STEM) education have been postulated to have positive impacts on Computational Thinking (CT) development. This study aims to discuss how these activities in primary school STEM education should be designed to nurture students with CT abilities. According to Sullivan and Heffernan (2016), CT learning progression with robotics consists of four stages, namely sequencing, causal inference, conditional reasoning, and systems thinking. Three examples about auto-piloting a robotic car to 1) run a square, 2) run along the white track, and 3) slow down and stop when it detects an obstacle are designed to illustrate the learning progression. The first example provides opportunities for students to develop their abilities of sequencing and causal inference. The second example demonstrates how conditional reasoning can be developed. The third example shows how systems thinking can be established. Based on this learning progression for CT development, an outline of STEM education with programmable robotics activities in formal and non-formal learning in primary schools is proposed. The key is that problem-solving should be the core of these STEM activities. Students’ knowledge in STEM related subjects in primary schools like science, mathematics and programming should be applied and in turn being consolidated.

Keywords: Computational thinking, primary schools, programming, robotics, STEM education

1. Introduction

Science, technology, engineering and mathematics (STEM) education has become increasingly popular in recent years. The reasons of such popularity are that STEM relates to a wide range of professions practiced in the physical world (Weintrop et al., 2016) and it can develop student’s realistic view for pursuing careers in the future (Committee on Prospering in the Global Economy of the 21st Century, 2007). Computational Thinking (CT), which is defined as “the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent” (Cuny, Snyder, and Wing, 2010, p.1), is closely related to STEM education. It is regarded as a core of all STEM activities and permeates all aspects of STEM (Basu et al., 2016). In addition, it acts as a consolidation of STEM experiences by deepening the understanding of STEM content areas for students (Sengupta, Kinnebrew, Basu, Biswas, and Clark, 2013). For example, STEM concepts like friction and environmental conservation are more understandable through engaging in programming and CT modelling activities (Hambrusch, Hoffmann, Korb, Haugan, and Hosking, 2009).

Although there are discussions about STEM and CT education (e.g., Sengupta, 2011; Sullivan and Heffernan, 2016; Weintrop et al., 2016), there have been few attempts to discuss how to conduct STEM education for CT development in primary school education. In particular, most of the previous studies have not emphasized how students might benefit from working with programmable robotics activities in primary school STEM education. Robotics focuses on physical object constructions, which could enhance students’ incentives in learning STEM-related concepts, confidence, interest and academic achievements (Park, 2015). Following the CT learning progression with robotics suggested by Sullivan and Heffernan (2016), this study illustrates some examples of programmable robotics...
activities that primary schools could be designed. Based on this learning progression, an outline of STEM education using programmable robotics for CT development in formal or non-formal learning in primary schools is proposed for nurturing the next generation to become creative problem-solvers.

2. Computational Thinking in STEM Education

Sullivan and Heffernan (2016) discussed how robotic construction kits (RCKs) operate as computational manipulatives in P-12 STEM education after an extensive review of literature. CT could be developed in STEM education through robotics activities, which engage students with hands-on experience by building a model, interact with the robot through coding. Based on the data collected from environmental sensors such as ultrasonic sensor, line follower sensor, or input modules like switches and buttons, the robots will function automatically according to the instructions set in the computational environment. They would react via output modules such as wheels, motor and buzzers.

Students might use RCKs such as LEGO Mindstorms and mBot for conducting robotics activities. In the process of using RCKs for solving STEM related problems, students can work with these computational manipulatives. Therefore, Sullivan and Heffernan (2016) proposed a four-stage learning progression in educational robotics activities for CT development. It is expected that students would acquire and demonstrate ability of the previous stage before moving to the next. Students could go through a progression from fundamental sequencing ability to reasoning abilities, including causal inference and conditional reasoning. The highest level of this progression is that students’ systems thinking will be improved. Table 1 shows the detailed explanations of each stage. By engaging in the robotics activities, students could also enhance their problem-solving skills by elementary approach like trial and error to more complex modeling method (Sullivan and Heffernan, 2016).

<table>
<thead>
<tr>
<th>Stage</th>
<th>CT Progression</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequencing</td>
<td>- Ability to put items into specific order</td>
</tr>
</tbody>
</table>
| 2     | Causal inference | - Comparison between the expected movement and immediate feedback from the programmed object  
|       |                 | - Hypothesis on why expected movement was not observed |
| 3     | Conditional reasoning | - Abstraction of a rule for the behavior  
|       |                 | - Use of environmental sensors to work with robotic device such as sensory-reason-action loop |
| 4     | Systems thinking | - Understanding of interacting functions of related parts of robotic device  
|       |                 | - Interaction between sensors, microcomputer and actuators such as motor and bulbs |

3. Examples Illustrating CT Progression Development through Programmable Robotics in STEM Education

The programmable robotics activities illustrated in this section were designed as group learning activities for Primary 5 students. These learning activities were successfully piloted in two primary schools in Hong Kong, in which six student groups were divided in each participating Primary 5 class with around 30 students for experiencing the learning progression of sequencing, causal inference, conditional reasoning, and systems thinking in the process of CT development.

3.1 Sequencing

Sequencing means the ability to arrange items into specific order to accomplish tasks with specific intention (Sullivan and Heffernan, 2016). Primary school students could develop their sequencing
ability by arranging blocks into specific orders in programming environment. For example, students can be asked to develop a program to auto-pilot a robotic car to run in a square as shown in Figure 1(a).

![Figure 1(a)](image)

Figure 1(a). An example that illustrates the Sequencing ability of students by auto-piloting a robotic car to run a square with codes in a generalizable pattern.

To accomplish this task, students have to remote control the car to run in a square. The following four steps in sequential order are needed: Step 1, run forward for a certain period of time (remark: this is the length of the side of a square); Step 2, stop running for a short period of time (remark: the car takes time to stop properly); Step 3, turn right for a period of time with a certain speed so that the car can turn right for 90 degrees from the original direction; and Step 4, stop running (remark: the car has to stop properly). The robotic car will run in a square and return to its initial position if the above sequences of steps are repeated for four times. However, when students start producing codes to run in a square, lines of codes as shown in Figure 1(b) are commonly found. The car might run forward with some steps and then more steps forward to complete running along one side of the square. The turning to right for 90 degrees procedures might be decomposed into turning to the right for 30 degrees and then 60 degrees. There is a need for teachers to guide students to learn from the stage with codes not ready to generalize a pattern to the stage of producing codes in running a square with pattern as shown in Figure 1(a).

3.2 Causal Inference

When students arrange the steps in order, they will run and test the programs at the same time. Causal inference will take place in this process. After comparing the actual movement with the intended movement of the robotic car, students would hypothesize the cause of such discrepancies. Using the same example in section 3.1 (piloting a robotic car to run in a square), causal inference is demonstrated in Figure 2.

![Figure 2](image)

Figure 2. An example illustrating the ability of students in causal inference by auto-piloting a robotic car to right turn 90 degrees.

Students might discover that the robotic car fails to turn 90 degrees as expected (Step 2 in Figure 2). Such discrepancy might make the robotic car fail to run in a square (Step 3 in Figure 2). When
the actual movement is not in line with the expected outcome, students may have different interpretations of the discrepancies. Students who demonstrate the ability of causal inference will be able to identify the fault sources and rectify them. In this case, the speed value of right turn and the duration of this process in the program have to be adjusted in order to find out the accurate parameters for turning 90 degrees in this case. After iterative inference and modification of the program, students might find out a possible solution. This experience can help students to develop the ability of causal inference.

### 3.3 Conditional Reasoning

Conditional reasoning refers to the ability to abstract rules to confine behavior in an environment. It involves logical reasoning in a programmable environment using data collected from sensors for defining proper operation of devices in the environment. Robots will follow the sensory-reason-action-loop computation. In other words, they would react to computation results based on the sensor input and take actions accordingly. In this example, students are asked to develop a program to make the robotic car move forward along the white track (see Figure 3(a)). In this task, results of the line follower sensors are imported to port 2 of the robotic car.

Figure 3(b) shows the value returned by the two-line follower sensors. Based on the sensor inputs, that is the values returned by the line follower sensors, students need to assign appropriate actions to the inputs. If the robotic car is not on the track and with the returned value of 0, the robotic car is programmed to stop. If the robotic car is touching the black edge on the left and with the returned value of 1, the robotic car is programmed to turn right. If the robotic car is touching the black edge on the right and with the returned value of 2, the robotic car is programmed to turn left. If the robotic car is exactly on the white track and with the returned value of 3, the robotic car is programmed to move forward. In this example, a sensory-reason-action-loop is constructed using multiple if-then conditionals.

![Figure 3(a). An example of preparing a coding table for conditional reasoning in piloting a robotic car to run along the white track.](image)

<table>
<thead>
<tr>
<th>Left sensor color</th>
<th>Right sensor color</th>
<th>Returned value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>Black</td>
<td>0</td>
</tr>
<tr>
<td>Black</td>
<td>White</td>
<td>1</td>
</tr>
<tr>
<td>White</td>
<td>Black</td>
<td>2</td>
</tr>
<tr>
<td>White</td>
<td>White</td>
<td>3</td>
</tr>
</tbody>
</table>

![Figure 3(b). An example of a piece of codes illustrating implementation of conditional reasoning in piloting a robotic car to run along the white track.](image)

### 3.4 Systems Thinking

Systems thinking refers to “judgement and decision making, systems analysis, evaluation and abstract reasoning” about the ways that different elements interact in a system (Sardone, 2017, p. 34). It offers a bird eye’s view of looking at connections in the world dynamically (Breil, Ritchie, and Greer, 2017). In the context of robotics, systems thinking refer to the understanding of functions, hierarchy, relationship, and interaction among sensors, microcomputers, actuators such as motor and bulbs, and other related components in the systems. The example in section 3.3 is further discussed in this section to illustrate the ability of systems thinking. Figure 4 demonstrates the adding of a program requirement to the task. The addition tasks of the robotic car are to slow down when an obstacle is detected afar and stop when an obstacle is nearby along the white track. Ultrasonic sensor is imported to port 3 of the robotic car to attain this goal.

In this example, an additional set of codes is added to the system. When the returned value of the line follower sensors is equal to 3 (i.e. the robotic car is moving along the white track), then a further
checking of the two conditions is needed. First, if the robotic car detects an obstacle in less than 50 units from itself, it will slow down (see codes in section A of Figure 4). Second, if the robotic car further detects an obstacle in less than 20 units, it will stop to avoid crashing with the obstacle (see codes in section B of Figure 4). In this example, the task requires students to understand interactions among the line follower sensors, ultrasonic sensor, wheels and motor (speed) of the robotic car, and the microcomputer (Arduino in this example) of the system in order to program the robotic car to function appropriately. This example illustrates that students possess an understanding of the hierarchy and the relationship of different robotic elements if they are able to construct a program to accomplish the tasks and these are the core elements to indicate students’ systems thinking ability.

Figure 4. An example illustrating the ability of systems thinking of students in auto-piloting a robotic car to move along the white track and additionally need to slow down and stop when an obstacle is detected.

4. An Outline of STEM Education Using Programmable Robotics Activities for CT Development in Formal and Non-Formal Learning in Primary Schools

In a bid to address STEM education for CT development in primary schools, this study proposes an outline of STEM education using programmable robotics activities. There are two possible opportunities for primary school students to develop CT in STEM education. One way is to incorporate STEM activities into formal curriculum subject learning such as science, mathematics, and programming in primary school curriculum. Another way is to promote STEM activities in non-formal learning such as co-curricular activities. Following Sullivan and Heffernan’s framework (2016) of robotics learning progression, Figure 5 shows the proposed outline.

![Diagram of the proposed outline](image)

Figure 5. An outline of STEM education using programmable robotics activities for CT development in formal and non-formal learning in primary schools.

The key is that problem-solving should be the core of these STEM activities. Students’ knowledge in STEM related subjects in primary schools like science, mathematics and programming should be applied and in return being consolidated. Through engaging students’ in problem-solving activities such as auto-piloting a robotic car in formal or non-formal programmable robotics doings, students’ knowledge in science such as friction, mathematics like relationship between speed and distance, and programming concepts of conditionals and loops need to be applied. These concepts
would probably in return being consolidated through these STEM activities. Students are expected to follow the learning progression of sequencing, causal inference, conditional reasoning, and systems thinking in the process of CT development. Experiences tell that students rarely program correctly at the first instance. In general, they need rounds of attempts such as adjusting the speed and direction of turning of the robotic car before they can program the car to run according to the instructions. Thus, students might demonstrate their abilities of sequencing, causal inference and conditional reasoning iteratively and interactively in the learning progression. Students might have the idea of systems thinking when they have experience of accomplishing the tasks of running along the white track and avoiding clashes with obstacles along the way. Students then can be encouraged to think creatively on how to improve further the functions of the robotic cars if more sensors can be introduced. This is a way to nurture them to become creative problem-solvers through this CT development process.

The goal of STEM education through programmable robotics activities is CT development. Examples in this study illustrate how STEM education in the context of solving robotics related problems in primary schools can be linked with the existing primary school curriculum for supporting the development of CT among primary school students. The robotics activities can provide opportunities for students to enhance confidence in applying STEM-related subject knowledge into solving real life problems in robotics context. Further work is to implement these activities in school formal or non-formal learning and conduct evaluation study on CT development.

Acknowledgements

The authors would like to acknowledge the support of the coolthink@JC project funded by the Hong Kong Jockey Club Charities Trust.

References


Designing Mobile Applications for Improving Positive Behaviour for Learning (PB4L) Pedagogy

Lou REDDY\textsuperscript{a,b}, Nilufar BAGHAEI\textsuperscript{a}, Gerhard VERMEULEN\textsuperscript{b}, Craig HILTON\textsuperscript{a} & Gregor STEINHORN\textsuperscript{a}

\textsuperscript{a}Unitec Institute of Technology, Private Bag 92025, Auckland 1142, New Zealand\textsuperscript{b}Wesley Intermediate School, Mt Roskill, Auckland 1041, New Zealand
lreddy@wesleyintermediate.school.nz

Abstract: Applications and online digital games are currently being used in New Zealand schools for the teaching of reading, writing and maths regularly. However, the teaching of behaviour has been heavily reliant on paper based and traditional forms of reinforcement as shown by the Positive Behaviour for Learning–School Wide (PB4L-SW) pedagogy. In this paper, we will introduce our current school based research for developing a tool for implementing PB4L-SW. We also describe our efforts in gamifying the teaching and reinforcing of behaviour through the design and development of the proposed “mPB4L” mobile application. We then outline our future plan for evaluating the effectiveness of the app in teaching Positive Behaviour and seeking teachers’ feedback at an Auckland-based Intermediate school. The findings will be used to support further development of integrating the PB4L into the digital world.

Keywords: gamification, behaviour management, game design, PB4L, teaching behaviour.

1. Introduction

Since 1989, the New Zealand education system has been largely punitive in its approach to behaviour management based on the reforms of the “Tomorrows Schools”. The New Zealand government at the time made the Ministry of Education have less administrative management of schools (Wylie, 2009). The Tomorrows Schools initiative was meant to decrease Maori (the indigenous people of New Zealand) under achievement and stand downs. A stand down is when a teacher decides that a student is to be removed from school for up to 10 days within a school year. Suspensions are considered a punitive approach and support little change in students positive behaviour (Lewis, Sugai, & Colvin, 1998). In 2015 - Hillier’s study of stand downs in New Zealand indicated that there were 14,198 stand-down cases (Hillier, 2015). This is a problem for educators as each of those cases of stand downs is when a student is not engaging with their learning and summarily not achieving. Maori and Pasifika learners therefore are disproportionately represented in the Hillier’s data due the behaviour management system set up to help them. The context of the research project is predominantly focused on Maori and Pasifika populations and therefore the forementioned data is pertinent to the study. Changes to practice could yield in positive outcomes for Maori and Pasifika learners however teacher responses to students behaviour needs to move away from the punitive responses in order to keep students highly engaged and in school (Guo wt al., 2011).

The problems associated with stand downs within New Zealand and its relationships with current behaviour management and instructional approaches have been a major concern of politicians and educators alike. In Tomorrows Schools, the original aim was to create more individualised and culturally responsive approaches to behaviour management, which never
eventuated. However in 2009, the New Zealand government’s Ministry of Education cited the Taumata Whanonga Behaviour Summit, which highlighted the need of addressing the disproportionate numbers of Maori learners being represented in stand down and under achievement data. The summit provided a catalyst for the changes in how behaviour management support via professional learning and development was offered to Schools (Ministry of Education, 2015). This paper gives a proposed intervention to address one of the gaps in the behaviour systems that support Maori and Pasifika learners.

A perennial issue within the New Zealand education sector is to improve the behaviour management approaches and to mitigate the shocking stand-down statistics highlighted by Hillier and the Ministry of Education (Boyd & Felgate, 2015). Studies on effective behaviour management approaches have shown their effectiveness in reducing stand downs and increasing learning outcomes for students (Lewis et al., 1998; Simonsen, Fairbanks, Bresch, Myers, & Sugai, 2008; Yeung, Mooney, Barker, & Dobia, 2009). This paper also acknowledges the ecological reasons for students’ misbehaviour that lead to stand-downs and proposes an instructional approach to behaviour management (Paquette & Ryan, 2001). There are a number of different approaches to support teachers in implementing non-punitive pedagogy and we will propose an adaptation to one such instructional approach.

Positive Behaviour for Learning School Wide (PB4L or PB4LSW) is an instructional approach to effective management of students’ behaviour within the whole school environment (Elder & Prochnow, 2016). The instructional approaches can be managed by teachers within the school setting. The strategies or tools used by educators to improve their pedagogical approach regarding classroom management should improve learner’s outcomes (Guo et al., 2011). The interventions also reduce student stand-downs and helps to create systems and procedures for effectively managing maladaptive or antisocial student behaviour (Lewis & Sugai, 1999). We use literature to make links to the practices of PB4L and aim to apply them into a digital online gamified learning platform by rewarding the prosocial behaviour (Bradshaw, 2013).

Wesley Intermediate School has implemented the PB4L-SW framework within the school since 2014. The main research question we are investigating in this project is whether a mobile-based PB4L tool is more effective than traditional methods of acknowledging students’ behaviour based on improving self-reported and user engagement data.

2. Related Work

The Ministry of Education established PB4LSW support for schools as one of the interventions adopted from the American system called Positive Behavioural Interventions and Supports (PBIS) (Elder & Prochnow, 2016; Lewis & Sugai, 1999). The set up of PB4LSW in New Zealand is based on the work by two key theorists – George Sugai from the Centre for Behavioural Education & Research, University of Connecticut and Tim Lewis who is the Dean for Research & Graduate Studies at the University of Missouri-Columbia (Ministry of Education, 2015).

Studies into the implementation of PB4LSW by the New Zealand Centre for Educational Research states that PB4LSW:

“offers primary, intermediate, and secondary schools a way of building a consistent and positive school-wide climate to support learning. It is a framework of key elements which schools implement in ways that suit their context.” (Boyd & Felgate, 2015).

PB4LSW systems is targeted at the continuum of procedures for encouraging expected behaviours or reward systems. The rationale for investigating the research around this area is contextualised to the implementation factors of PB4LSW at Wesley Intermediate School since 2015. Theorists suggests that teachers should use reward tools and strategies to help students improve their behaviour, with a goal to reinforce the expected behaviour within a consistent
reward system and school wide (Johansen, Little, & Akin-Little, 2011; Parsonson, 2012; Simonsen et al., 2008).

As part of ongoing evaluation Wesley Intermediate School uses the School wide Evaluation Tool or SET (Horner et al., 2004) to collect data on the effective implementation of PB4L-SW systems. The school initially had lower levels of teaching expected behaviours at the start of training. We have slowly improved our systems and also in a decrease in our stand downs since implementation. See Figure 1 of SET from 2013 to 2016 (Wesley Intermediate School, 2016). As reflected in our SET survey, there is a lack of PB4L expectations being taught and acknowledgement systems being embedded. The evaluation of PB4L –SW similarly shares evidence that implementation PB4L systems in intermediate and high schools need more development compared to Primary schools within the study (Boyd & Felgate, 2015). The review of the current data teacher feedback showed that teachers gave a higher priority to reward systems for non-classroom settings whilst using an Effective Behaviour Survey which is similar to user feedback surveys (Lewis & Sugai, 1999; Ministry of Education, 2017a). Teachers at the school wanted to improve the traditional paper based tokens as rewards for their students. Some educators still use traditional paper based systems as simply they do not have the knowledge or technological skills to create digital tools (Naismith, Lonsdale, Vavoula, & Sharples, 2004). Research shows that teacher’s perspectives towards implementing behavioural interventions is heavily influenced by their understanding or prior education (Johansen et al., 2011).

Research into the implementation of PB4L systems in New Zealand show the need for more development in the engagement of teachers and students in the intermediate and high school sectors (Boyd and Felgate, 2015). In Boyd’s research most primary school coaches (managers) expressed that their staff effectively used acknowledgement and behaviour consequence systems to encourage positive behaviour, however, a quarter (24%) of secondary/intermediate coaches disagreed with the primary school coaches’ statement (Boyd and Felgate, 2015). Educators needs more tools, yet there are not many evidenced based tools for educators to use, so the proposed interventions aim to design a digital tool to support teachers trying to implement PB4L pedagogy in their school (Naismith et al., 2004).

3. Our Proposed mPB4L App

Our proposed app, mPB4L, is based within the context of a NZ Intermediate School’s behaviour management system and inherent in this system are teacher’s perceptions and

![Figure 1. SET survey 2016 Wesley Intermediate School](image-url)
efficacy of PB4L. The design and implementation of mPB4L app is making the learning of pro-social behaviour gamified and more engaging for teachers and learners (Schoech, Boyas, Black, & Elias-Lambert, 2013).

We will trial the mPB4L app later this year and evaluate its effectiveness in supporting teachers to implement PB4L pedagogy. As a group of leaders and researchers of PB4L, we have found that there is relevant research promoting gamified forms of digital education to help teachers and students achieve positive behaviour outcomes (Gouseti, 2014 and Kapp, 2012). This app will encourage learners and teachers to use the application competitively and also promote high expectations for Pasifika and Maori learners which will foster better relationships (Hawk, K., Tumama-Cowley, E., Hill, J., & Sutherland, 2002). The app will aim to gamify the teaching and learning of behaviour and the participants’ engagement with the app and enjoyment will be measured.

Figures 2-3 shows an initial prototype of the app interface for students with links to the acknowledgements that students may have received from their teachers. The concept heavily focused on students reinforcing each other’s PB4L values (respect others, yourself and the environment) within their profile page. Figure 4 shows the user dashboard for the teacher view of assigning points.

Some potential features could include the app being able to record incident report data, functional behavioural assessment (O’Neill & Stephenson, 2010) and response strategies for antisocial behaviour for teachers which is divergent to similar apps like Class Dojo. We believe our trial and findings will create some meaningful shifts in student wellbeing, decreasing stand-down rates and also increasing teacher engagement of PB4L in the future. The research before and after the trial will inform further design of the mPB4L app.

4. Conclusions & Future Work

In this paper, we outlined a proposed rationale for testing a prototype of the forementioned design and development of mPB4L app, based on evidenced based practice of PB4L (Ministry of Education, 2017b). We have also demonstrated a link to effective use of digital tools such as gamification within apps (Muntean, 2011; Simões, Redondo, & Vilas, 2013) and use of apps in education to provide accessibility and changing students behaviour (Naismith et al., 2004). The long term goal is to make an app that encompasses PB4L pedagogy and research teachers perceptions on whether it helps improve learning relationships and reduce stand down rates (Hillier, 2015).
The methodology will drive what the teachers want in their mPB4L app and we will use an indigenous approach to research the effectiveness of the app. The approach called Talanoa uses the method of informal story telling and will be used during two focus group sessions (Vaioleti, 2006). The Talanoa sessions will be spread over a term (first before intervention and after intervention). The approach is post-positivistic and encompasses teachers as practitioners conducting research to create solutions (Koshy, Valsa, & Waterman, 2010). The research will focus on whether teachers found the app beneficial in implementing PB4L Pedagogy. Participation is voluntary, not performance based and will have qualitative data collection through Talanoa and quantitative data collection through surveys like the Effective Behaviour Survey (Ministry of Education, 2017a), which is already in place. The Talanoa methodology will allow for culturally inclusive practice. A careful understanding of protocols, values and principles will be needed, for example the considerations made of the differences in male and female relationships, customs of individuals titles and roles (Pihama, Tiakiwai, & Southey, 2015; Vaioleti, 2006). We believe our findings will create some meaningful shifts in student wellbeing, decreasing stand down rates and increasing teacher engagement of PB4L in the future.

Acknowledgements

The lead author would like to thank all the people who revised previous versions of this document. The latest prototype concepts of the mPB4L application is designed in collaboration with Gerhard Vermeulen, Rebecca Reddy and Jenny Barker.

References


Teachers’ Perception of IT in Science Education

Der-Thanq CHEN* & Wenjin Vikki BO

* National Institute of Education, Singapore
b Singapore University of Social Sciences
*victor.chen@nie.edu.sg

Abstract: In this paper, we report findings from a qualitative inquiry into teachers’ perceptions and experiences of using IT, specifically simulations in science education. Semi-structured interviews were conducted with 12 science teachers. Thematic analysis reveals that teachers’ perception towards simulation use in science teaching included three aspects: (a) perceived affordance of teaching with simulation in the science classroom, (b) perceived affordance of learning with simulation in the science classroom, and (c) perceived affordance of simulation infrastructure in the science classroom. With the attitudes described above, most teachers mainly had the experience of adopting simulation for demonstration purpose in a teacher-led instruction. Attempts to let students operate simulation themselves and explore alternative modelling did not seem to work well, considering the inadequate school facilities and limited classroom time to cover the required syllabus. Based on the findings, we propose the FLIPPING framework – using interactive simulation in flip learning approach. We believe the 8 key components in FLIPPING framework (Flexible environment, Learning culture, Intentional content, Professional educators, Preparedness for learning, Infrastructural readiness, Novice-proof interface, and Guided pragmatism) can alleviate both material difficulties of simulation implementation (e.g. infrastructure), and non-material challenges (e.g. student readiness) and promote the use of interactive simulation in science teaching in sustainable ways.

Keywords: Science Education, Teacher Perception, IT

1. Introduction

ICT, specially simulations, has been increasingly influential in science education because it can help students visualize abstract scientific concepts (McElhaney & Linn, 2011; Wu & Huang, 2007). On one hand, by allowing students to visualize abstract concepts as well as to explore and test scientific modelling, computer simulations have been claimed to promote students’ deep learning and conceptual understanding in science, such as optical lenses (Chang et al., 2008), moon phases, trajectory motion, relative motion (Monaghan & Clement, 1999), etc. In addition to greater achievement, computer simulations were also found to significantly enhance students’ positive attitudes towards science subjects.

While a growing number of simulations have been developed, research has shown that simulation implementation in the science classroom is slow and challenging (Mehlinger & Powers, 2002; Pelgrum, 2001; Schwarz, Meyer, & Sharma, 2007, Schrum, 1999; Strudler & Wetzel, 1999). To solve those issues, most recent studies focused on increasing teacher training and developing pedagogical strategies to use interactive simulation in science classrooms (Khan, 2011; Kopcha, 2012; Schwarz & Gwekwerere, 2007). However, our early interactions with teachers indicated that the extent of simulation implementation still seemed to be quite low, despite the organized teacher training and developed pedagogical methods.

This paper reports a study to understand science teachers’ perceptions and experiences of simulation implementation after workshop training. Individual interviews with different teachers from various schools across Singapore were the main source of data which were later triangulated with their lesson plans and students’ assignments. The findings of the study will yield useful insights of the implementation process and will suggest an approach that could potentially support the implementation of interactive simulation in science education in sustainable ways.
Although it has been acknowledged that computer-based modelling using simulations can facilitate modelling-oriented instruction and students’ deep learning of physics concepts (Goh, Wee, Yip, Toh, & Lye, 2013; Wee et al., 2012), research on how teachers can effectively integrate simulations into physics classes is rather scarce in Singapore. The present project is a professional development and research study that examines teachers’ use of simulations after participating in hands-on workshops about the tools and related pedagogical approaches. This paper explores science (physics and chemistry) teachers’ perceptions of teaching with simulations after their training workshops. We also aimed at understanding how their perceptions shaped their simulation implementation in classrooms, and what challenges do they face during the implementation process.

In total, 12 teachers from 7 schools across Singapore agreed to participate in the individual interviews. Among the 7 schools, 5 were government schools where students’ proficiency levels varied and 2 were independent schools with high achieving students in Singapore.

2. Methodology

The goal of the present research was to explore science teachers’ post-training perceptions towards simulation use in class, as well as their teaching practice when implementing simulation in science classroom. A qualitative study (Merriam, 1998) was adopted for this study as it allowed for exploring the phenomenon through the perspectives and experiences of the subjects in the study.

The primary method of data collection for the teachers’ perspectives was individual semi-structured interviews. Among the 12 interviews, 11 were conducted face to face and 1 was via telephone. Semi-structured interview was conducted to allow for the variation in the question order as well as potentially additional questions and probes to particular individuals (Creswell, 2007).

Thematic analysis (Miles & Huberman, 1994) was used to abstract the data. Patterns within the data were developed and then categorized into codes and themes. To ensure the accuracy of interpretation in the qualitative data, other experienced scholars in qualitative research were invited to categorize themes based on the research questions and literature. Disparity in developing codes and themes was discussed until a complete agreement was reached. The consistent triangulation in the data analysis among various researchers prevented the potential risk that the interviews might be analysed only from one researcher’s own subjective standpoint (Yin, 2013). From the process of thematic analysis, two major themes appeared, including teachers’ perceptions of interactive simulation in science education (what they think after training), and teachers’ past experiences of simulation implementation (what they did after training) in the classroom practices.

3. Findings

We present findings on teachers’ perceptions of stimulation use in science teaching first, followed by their implementation experience in class.

3.1 Teachers’ Perceptions towards IT in Science Teaching

Although teachers agreed on how simulation can enable students to visualize abstract concepts in an interactive way, they tended to demonstrate more concerns than favour. Their concerns included three major aspects: (a) perceived affordance of teaching with simulation in the science classroom, (b) perceived affordance of learning with simulation in the science classroom, and (c) perceived affordance of simulation infrastructure in the science classroom.

Simulations afford visualization capacity for more abstract topics. All the 12 science teachers considered interactive simulation to be useful only for instructing certain abstract topics, such as kinematics, dynamics, magnetism, etc. They all admitted the “innovative and fun part of technology in teaching”, but the major advantage of simulation in science teaching seemed to be limited to the visualization of abstract concepts. Under such circumstances, teachers would consider using simulations in class based on the availability of resources provided as well as the match between the provided resources and their teaching plans of specific topics. Teachers with technical backgrounds in
ICT tools are also open to the customization of simulations while adopting them in teaching. Simulations do not afford benefits equally for all topics. More concrete and easily observed phenomena may be better introduced in a laboratory setting or even a traditional classroom, as described by interviewed teachers — either instead of or as a precursor to using simulations. When it comes to other topics, however, most teachers considered simulations to be redundant, especially when you have the option of demonstrating scientific concepts in real-life settings, either in classroom or laboratory. Furthermore, teachers find simulations are also limited to the more proficient students while using it in science teaching. Among the 12 interviewed teachers, 10 teachers demonstrated the concern of students being unable to catch up with the simulation in class. The exception though are the 2 teachers from the top schools in Singapore where students’ proficiency level are among the highest in the country. According to the teachers’ interviews, teaching with simulations seemed to confuse many of their students, especially the weaker ones. Therefore, they felt they had “no choice but to revert back to pen and paper”.

After receiving trainings of interactive simulation and trying it in class, teachers discovered mixed effects on student learning. On one hand, it was consensus among the majority that students became much more engaged in learning when simulation was used to demonstrate abstract phenomenon in science. Different from the traditional approach when students would just sit down, listen to the teacher and look at the textbook, simulation was believed to be “a strong variation of learning approach”, which was effective to link different topics as well as promote students’ learning interest. On the other hand, some teachers noticed the complication simulations brought to learning in class, especially when the students were asked to operate the simulations themselves. This could be due to two major reasons: the need for extra learning of the software and potentially misleading representation in the simulation. In support of this statement, another teacher also from a government neighborhood school mentioned that most of the students from neighborhood schools are not the academically inclined students in Singapore, and “they might not be much appreciative of the new technology in learning”, not to mention they need to learn extra just to use the technology. Notably, despite the learning engagement or learning complication simulation has brought to students, the teachers’ biggest concern is about whether simulation will really benefit students’ learning outcome, especially in examinations. One teacher pointed out that what the students finally need to do is a paper-pen test where is no simulation. Hence they prefer to continue with learning in traditional ways such as writing down and calculating, etc. Another teacher agreed and added that within the limited time for the students to learn, he tended to give up using simulation and maintained the conventional approach because “ultimately we still need to get them to be prepared for examinations”.

Teachers identified a variety of challenges in school infrastructure which they thought constrained the implementation of simulation in science teaching. First and foremost, students do not always have individual access to computers when needed. So teachers in government neighborhood schools struggled to implement technology in class due to the shortage of school facilities to begin with. Internet bandwidth is another challenge that most government neighborhood schools encountered since interactive simulation required fast and stable network while school internet barely reached the requirement. Even though independent schools might have eliminated those two challenges in hardware, the limited materials in simulation software were still restricting the potential adoption in class, especially the materials matching the context of the Singapore science curriculum. Without overcoming the challenges both in hardware and software, teachers would often doubt the possibility to implement simulation in science class, even many of them might have realized the potential benefits simulation can bring to science teaching and learning.

3.2 Teachers’ Experience of Using IT in Science Teaching

With the attitudes described above, most teachers mainly had the experience of adopting simulation for demonstration purpose in a teacher-led instruction. Attempts to let students operate simulation themselves and explore alternative modelling did not seem to work well, considering the inadequate school facilities and limited classroom time to cover the required syllabus.

According to self-report accounts during interviews of how the teachers attempted to use simulations in their teaching, 10 out of 12 teachers used direct instruction while adopting simulation in their classes. The implementation was mainly for demonstrating scientific concepts without involving
student interaction. They explained that they had limited classroom time to complete the school syllabus and prepare students for their final exams and so they could not afford to make the class more student-centered and provide opportunities for their students to operate the simulations themselves in class. Some teachers reported designing inquiry-based instructions occasionally in computer labs after they have completed the required syllabus. However, only the more proficient students seemed to be engaged while the less proficient students found difficulty catching up and hence lost interest in the lesson. As Teacher 11 described, the less proficient students in her class had no idea what the aim of the inquiry-based activity was, and they would always tend to look for answers, asking “what exactly do you want me to look out for? What do I have to do? Can you tell me what to do?” As a result, the lesson became more teacher-centered employed direction instruction in the end. In contrast however, the two teachers from the top schools used inquiry-based instructions a lot more while implementing interactive simulation in class. Admittedly, students were expecting more teacher guidance in such activities, but no obvious difficulty was indicated when they were asked to use the simulations and explored the alternative modelling under minimal teacher guidance. It was obvious that in most of the schools where students’ proficiency level was average, both the teachers and students preferred direct instruction while using simulations in class. This enabled the teachers not only to fulfil their teaching duties within limited classroom time but also to ensure efficient students’ knowledge acquisition and exam preparation. In contrast, teachers in the top schools were more willing to adopt inquiry-based instruction which their students were able to follow.

All the teachers attempted to integrate simulation in science teaching and learning after class, either as a required homework or an optional exploration for students to use simulation in their spare time. In this way, teachers felt they “did not have to distribute the limited classroom time to the use of simulation”, and students were also enabled to explore the alternative modelling at their own pace. Notably, students were found to be responsive to learning with simulation after class only when it was required as homework, and tended to ignore the exercises when it was optional. Furthermore, certain amount of teacher guidance was also needed when students conducted self-learning after class as they needed help with not only the content knowledge of the subject, but also the technological knowledge to operate the software. As can be seen, teachers in our study seemed to embrace simulation use after class especially to promote the inquiry-based learning among students at different proficiency levels. The use of simulations after class also eased the facility shortage in school since students would have more places and platforms to use simulation in their spare time.

4. Conclusion

The findings in this study suggest that despite teacher training and technical support, the implementation of interaction simulation was perceived to be difficult among science teachers in Singapore context. Although certain teachers attempted to use simulation in their class, future intent to continue using was indicated to be low due to the limited classroom time to complete the required syllabus, and the pressing concern whether the use of interactive simulation can directly contribute to exam results. For some schools that were not well equipped with technological infrastructure, simulation implementation was impossible to begin with regardless of how positive teachers perceived the use of simulation in class.

Faced with the challenges both material (e.g. infrastructure, technical support) and non-material (e.g. teacher training) in implementing simulations in science education, most of the current literature focused on creating instructional models and strategies in classroom teaching (Khan, 2011; Schwarz & Gwekeberere, 2007). However, based on the findings of the present study we propose a more sustainable approach: flipping learning with interactive simulation. We extend the FLIP model (Hamdan, McKnight, McKnight, & Arfstrom, 2013) and FLIPPED model in higher education (Chen, Wang, & Chen, 2014) by proposing the FLIPPING framework with eight key components: Flexible environment, Learning culture, Intentional content, Professional educators, Preparedness for learning, Infrastructural readiness, Novice-proof interface, and Guided pragmatism. Informed by the present findings, this extension would better fit science education in Singapore context.
References


Khan, S. (2008). What if scenarios for testing student models in chemistry Model based learning and instruction in science (pp. 139-150): Springer.


Surveying Indonesian Teachers’ Design Belief and TPACK for 21st Century Oriented Learning

Ching Sing CHAI*, Joyce Hwee Ling KOHa, Uma NATARAJANb, Pei-Shan TSAIc*, Murni RAMLId & Ari WIDODOe

aNational Institute of Education, Nanyang Technological University, Singapore
bHead Foundation, Singapore
cTeacher Education Center, National Taipei University of Technology, Taiwan
dFaculty of Teacher Training and Education, Sebelas Maret University, Indonesia
eFakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia, Indonesia

* Sandra.pstsai@gmail.com

Abstract: This study validates a survey created to study Indonesia teachers’ beliefs about lesson design and their technological pedagogical content knowledge (TPACK) for 21st century oriented learning. The teacher’s design beliefs, which is a new perspective in understanding teachers’ beliefs, consist of their view about their design disposition (DD), beliefs about the new culture of learning (BNCL), and their beliefs about teacher’s role as designer (TAD). The TPACK survey, which is also newly created, examines the teachers’ efficacy about their knowledge of using technology for active and constructive learning, authentic learning and collaborative learning instead of adopting the seven factors TPACK model. Exploratory factor analysis and reliability analysis reveal that the six factors survey is valid and reliable. In addition, all factors are significantly correlated. Path analysis reveals that two of the teachers’ design beliefs (DD and BNCL) predict the teachers TPACK for 21st century oriented learning. The results imply that it is likely necessary to consider teachers’ design beliefs when teacher educators plan to foster teachers’ TPACK.

Keywords: Technological Pedagogical Content Knowledge, Teachers’ beliefs, Professional development

1. Introduction

Twenty first century oriented learning is undergirded by the meaningful use of technology to engage students in solving authentic real-world problems through active/constructive learning in collaborative classroom (Dede, 2010; Howland, Jonassen & Marra, 2014; Voogt & Roblin, 2012). To engender such learning in today’s classroom demands teachers to assume the role of designer to psychologize the subject matter as real world problems that require the use of ICT as cognitive tool to help in resolving (Howland et al., 2014). Teachers’ professional development and the school ICT provision, teachers’ beliefs and their design capacity, however, have been identified as three level of interacting barriers that may hinder the actualization of 21st century oriented learning with ICT (Tsai & Chai, 2012; Voogt, Erstad, Dede, & Mishra, 2013). A current direction to help to address the problem of developing teachers as designer is to develop teachers’ technological pedagogical content knowledge (TPACK) through learning by design (Howland et al., 2014; Voogt et al., 2013). To facilitate effort in this direction, it seems appropriate to develop instrument that can validly measure the relevant aspect of teachers’ beliefs and their 21st century oriented TPACK.

This study validates a survey created to study Indonesia teachers’ beliefs about lesson design and their technological pedagogical content knowledge (TPACK) for 21st century oriented learning. The teacher’s design beliefs, which is a new perspective in understanding teachers’ beliefs, consist of
their view about their design disposition (DD), beliefs about the new culture of learning (BNCL), and their beliefs about teacher’s role as designer (TAD) (see Koh, Chai, Tsai, 2015; Chai…2017; Chai & Koh, in press). The TPACK survey, which is also newly created, examines the teachers’ efficacy about their knowledge of using technology for active and constructive learning, authentic learning and collaborative learning. These dimension of learning are essential for 21st century learning (Howland, Jonassen & Marra, 2014; Voogt & Roblin, 2012). Further reviews of current research about teachers’ beliefs and TPACK survey are elaborated in the following paragraphs.

1.1 Studies of Teachers’ Beliefs

Teachers’ beliefs are multidimensional in nature. Past research on teachers’ beliefs can include their epistemological beliefs, pedagogical beliefs, beliefs about subject matter; which in the context of ICT integration have been identified as influencing teachers’ decision (Ertmer, Ottenbreit-Leftwich, & Tondeur, 2014). Given the current concern about teachers’ design capacity, Chai, Tan, Deng and Koh (2017) and Koh, Chai, Hong and Tsai (2015) have identified some facets of teachers’ beliefs including their design disposition, beliefs about new culture of learning (see Thomas & Brown, 2011) and their beliefs about themselves as designers (Chai & Koh, in press). The cumulative research efforts of this group of researchers have gradually build a constellation of teachers’ design beliefs that have reportedly influence the teachers’ design work. However, studies of teachers’ design beliefs are just emerging, and it should be tested in a wider context to examine its effect, especially on how it relates to the teachers’ TPACK. This study adopted the survey design by Chai and Koh (in press) to study whether or not such beliefs can be identified among the Indonesian teachers.

1.2 Current Development of TPACK Survey

Previous TPACK surveys generally adopt the seven factors model (Chai, Koh & Tsai, 2016; Mishra & Koehler, 2006). While Chai et al.’s (2016) review indicates that earlier research generally fails to establish the seven factor model (see e.g. Archambault & Crippen, 2010), Chai et al. (2016) have concluded that most current survey are able to measure the seven factors model validly. Nonetheless, Chai et al. (2016) suggest that each factor can be further expanded to include distinguishable sub-factors. For example, Yeh, Hsu, Wu, Hwang, and Lin (2014) has adopted the Delphi method to identified an alternative TPACK-practical framework that generated eight different dimensions of TPACK for Taiwanese science teachers. In other words, TPACK factors can be perceived differently under different sociocultural and pedagogical contexts. This study attempts to focus on the final TPACK factor that synthesizes teachers’ technological knowledge, pedagogical knowledge, content knowledge and its overlapping constructs (i.e. technological content knowledge, pedagogical content knowledge and technological pedagogical knowledge). It does not adopt the seven factors model. Instead, it hypothesized that the final TPACK can be further factorized with different pedagogical emphasis denoting more refined form of knowledge that teachers may need to master to design and implement ICT-oriented 21st century learning. Consequently, three distinctive yet commonly emphasized dimensions of 21st century learning were used to expand the TPACK factors (see instrument section).

Given the review above, this study attempts to answer the following research questions:

1. Is the instrument valid and reliable for the purpose of surveying Indonesian teachers’ design beliefs and their 21st century learning oriented (21CLO) TPACK?
2. What are the relationships between the Indonesian teachers’ design beliefs and their 21st century learning oriented TPACK?

2. Methods

2.1 Participants and Survey Instrument
There are 187 participants in total who have volunteered for this study. These teachers have different profiles: preservice K-12 teachers (54), practicing K-12 teachers (66) and university teachers (67). The teachers are from Solo, Jakarta and Bandung area. Sixty out of 187 are male teachers. Their mean age is 30.7 years (SD=9.5). Around 50.4% of them are from the mathematics, science or technology teachers while the rest are language, religious studies, and social studies teachers.

The survey instrument comprises of three parts: demographic, teachers’ design beliefs and the 21st century learning oriented TPACK (21CLO-TPACK). Demographic data collected include age, gender, teaching level and subject matter taught. The teachers’ design beliefs comprise three scale namely design disposition (DD), beliefs about the new culture of learning (BNCL), and their beliefs about teachers’ role as designer (TAD). DD refers to the teachers’ assessment of their comfort level about ill-defined situations that calls for design thinking to resolve. BNCL refers to the teachers’ beliefs about whether today’s learners should be engaged in creating digital artifacts as a means to learn. TAD refers to the teachers’ acceptance of their role as designer for 21st century classrooms. The survey is adopted from a recently validated instrument (Chai & Koh, in press). The 21CLO-TPACK also comprises of three scales: TPACK for active and constructive learning (ACL), authentic learning (AUTL) and collaborative learning. The survey was presented in a seven-point Likert scale, ranging from 1 for strongly disagree to 7 for strongly agree. This is a newly constructed survey based on currently understanding of 21st century learning. All items were subjected to review by three education professors for the assessment of face validity.

2.2 Data Collection and Analysis

The participants were invited to participate in the survey through e-mail by the Indonesian co-authors of this paper. The survey is an English-version survey and it took around 15 minutes to complete. Indonesian co-authors and education professors have reviewed the survey to make sure that the Indonesian teachers can understand each survey item. The data were cleaned and subjected to exploratory factor analysis (EFA), Cronbach Alpha’s analysis and the means of the validated scales were computed. Pearson correlation was then conducted followed by path analysis.

3. Findings

3.1 Results of Exploratory Factor Analysis (EFA) and Reliabilities Analysis

Exploratory factor analysis employing principal axis factoring with direct Oblimin rotation was conducted. Six factors were extracted and items with factor loading greater than 0.5 were retained. Ten items with cross loading and insufficient factor loadings were removed. The overall Alpha reliability of the survey is 0.95. Table 1 shows the final survey items, and reports the outcome of EFA with factor loadings, alpha reliabilities, mean scores of the factors and the standard deviation. In addition, the survey indicates that the Indonesian teachers hold strong beliefs that teachers should be designers and students learning should be driven by creating digital artifacts in participatory culture. While they generally possess design disposition, their efficacies of 21CLO-TPACK should be enhanced.

Table 1: Outcome of EFA and Reliabilities Analysis

<table>
<thead>
<tr>
<th>TPACK Active and Constructive Learning (ACL) (α=0.91), Mean=5.15, SD=1.03</th>
<th>Factor Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I know how to choose appropriate technologies based on the topics I am teaching for students to perform student-centered inquiry.</td>
<td>.894</td>
</tr>
<tr>
<td>2. I am competent in helping my students to critically synthesize information from various web-based resources for content learning.</td>
<td>.800</td>
</tr>
<tr>
<td>3. I can facilitate students’ ongoing effort in designing solutions with technology using their content knowledge.</td>
<td>.773</td>
</tr>
<tr>
<td>4. I am able to use technology to stimulate my students’ higher order thinking about the subject matter.</td>
<td>.739</td>
</tr>
</tbody>
</table>
TPACK Authentic Learning (AUTL) (α=0.92), Mean=5.04, SD=1.09
1. I can arouse students’ interest in solving real world problems using subject related software. .889
2. I am competent in searching for online video resources to initiate real world problem solving related to the subject matter. .722
3. I can use technologies to scaffold students’ in solving complex problems arising from the topics that I teach. .707
4. I can engage students in learning the subject matter using the ICT tools that subject matter experts use. .660

TPACK Collaborative Learning (COL) (α=0.93), Mean=4.81, SD=1.13
1. I can engage students in substantial peer critiquing work through collaborative software. .883
2. I am competent in prompting students to talk deeply about the content knowledge in online platforms. .856
3. I can facilitate students’ co-construction of subject matter representations when they are working in small groups around a computer. .854
4. I can formulate in-depth discussion topics about the content knowledge for students’ online discussion. .694

Design Disposition (DD) (α=0.87), Mean=5.36, SD=0.96
1. I am comfortable with occasional failures from trying out new approaches for teaching. .924
2. I am comfortable to explore conflicting ideas. .740
3. I am comfortable to deviate from established practices. .696
4. I am constantly seeking to turn constraints into opportunities. .648

Teacher as designer (TAD) (α=0.93), Mean=6.01, SD=0.96
1. Working like designer is part of the teacher’s duty. .928
2. Teachers should devote substantial time to design lessons. .907
3. It is my responsibility to master the skills of designing lessons. .759

Beliefs about New Culture of Learning (BNCL) (α=0.91), Mean=5.93, SD=0.91
1. Students should be able to choose relevant digital resources for self-initiated learning .897
2. Remeshing digital resources responsibly is a good way to learn. .854
3. Today’s learners should be able to remix relevant resources to publish their ideas. .799
4. Managing personal online learning resources is a desirable skill. .797
5. Producing creative digital works is a meaningful task .761
6. Online collaboration should be part of students’ personal competencies. .737

3.2 Correlation and Path Analysis

Pearson correlations were generated between the six factors. In general, all six factors were significantly correlated with Pearson correlation coefficients ranging from 0.74 to 0.29. Table 2 below documents the correlations. To investigate the roles that teachers’ design beliefs in their 21CLO-TPACK, this study used the path analysis technique to examine the relationships among these variables. The teachers’ design beliefs factors were considered as predictors, while the 21CLO-TPACK factors were viewed as outcomes variables. The results revealed that several significant relations among the factors in teachers’ design beliefs and those in the 21CLO-TPACK, as shown in Figure 1. First of all, DD significantly explained all of the 21CLO-TPACK factors, including AUTL (β = 0.32, p < 0.001), COL (β = 0.33, p < 0.001), and ACL (β = 0.31, p < 0.001). The teachers who hold stronger beliefs about design disposition (DD) would possess stronger efficacy in AUTL, COL and ACL. In addition, BNCL played a positive role in AUTL (β = 0.15, p < 0.05) and ACL (β = 0.27, p < 0.001). This finding indicated that the teachers who are more inclined towards new culture of learning (BNCL) would have stronger efficacies for AUTL and ACL. Teacher as designer is however only positively correlated to the teachers’ 21CLO-TPACK without predictive relationships.

Table 2: Pearson Correlations

1004
This study aims to validate a survey instrument that is partly adopted (Chai & Koh, in press) and partly created to examine Indonesian teachers’ design beliefs and their 21CLO-TPACK. The findings that the survey possess factorial validity and it is reliable. The survey outcomes indicate that the Indonesian teachers are inclined towards designing instruction for 21st century learning mediated through participatory co-creation of digital artifacts (Thomas & Brown, 2011). The study therefore contributes first towards the cross-cultural validation of the teachers’ design beliefs. Teachers’ design beliefs are emerging to be an area of concern when teachers are engaged in designing technology-based instruction (Ertmer et al., 2014; Tsai & Chai, 2012) and the cross-cultural validation may path the ways for future cross-cultural research in this area. In addition, the establishment of the 21CLO-TPACK survey may contribute to more specific assessment of teachers’ efficacies to integrate technology for 21st century oriented learning (Chai et al., 2016). As indicated in the findings of this study, the Indonesian teachers possess strong beliefs that education with technology should move towards the new culture of learning and teachers should be the designer for such learning. Nonetheless, the teachers do not possess strong efficacies in designing 21st century learning with technology. The Indonesia education authority should devote time and effort to facilitate teacher professional development in this area. Perhaps scaffolding the teachers to learn by design supported by the scaffolded TPACK lesson design model as reported by Chai and Koh (in press) could be an effective way to enhance the teachers’ 21CLO-TPACK.

We have conducted several t-tests and ANOVA to further examine the results base on the demographic. However, there is generally a lack of significant difference between genders and the levels they teach. The only differences detected so far is that the mathematics, science and technology teachers (generally the hard disciplines) possess stronger AUTL then the language, social studies and religious studies teachers ($t = 2.11, p < 0.05$).

Further analysis of the relationships between the teachers’ design beliefs and their 21CLO-TPACK indicates that the Indonesian teachers’ design beliefs and their 21CLO-TPACK are correlated and the teachers’ design beliefs could generally predict their 21CLO-TPACK with the exception of TAD. This further imply that it could be important to consider the teachers’ design beliefs as these beliefs would shape their design. One possible way to address the issues of teachers’ design beliefs is to explicitly discuss the beliefs before and after the teachers’ design technology-based lessons. We conjecture that it could be helpful if the teachers could implement the designed technology-based lessons and discuss the outcomes.
There are some limitations to the current study. First, the study adopted convenience sampling which require the findings to be viewed with caution. Future research could adopt sampling strategies that are statistically more powerful. In addition, not all 21st century learning practices were included in the survey. For example, self-directed learning is also frequently mentioned in 21st century learning framework (Voogt & Roblin, 2012). Future research may consider including this dimension for the 21CLO-TPACK.

Acknowledgements

We would like to thank the Head foundation and all participating Indonesia Institutes of Education for supporting this research.

References


Supporting Learning by Doing in a Work-process-oriented Curriculum

Yongwu MIAO\textsuperscript{ab}\* & H. Ulrich HOPPE\textsuperscript{a}
\textsuperscript{a}Department of Computational and Cognitive Sciences, University of Duisburg-Essen, Germany
\textsuperscript{b}Wuhan E-competence Education LMT., Wuhan, China
\*miaoyw@yahoo.com

Abstract: Work-process-orientation is a new concept in vocational education and training (VET). Structure and content of a work-process-oriented curriculum (WPOC) are derived from a "typical professional task" in an occupation. This new curricular framework for VET is called "learning arena". This paper analyzes the characteristics of a learning arena and identifies technical requirements to support an online or blended WPOC. Based on these requirements we suggest an approach to systematically support learning by doing within a WPOC. The approach can be characterized as: guide the learning process according to the corresponding WPOC model, enable doing by integrating work-related application tools, support learning by embedding associated learning materials in application situations, and conduct outcome-based assessment by collecting and evaluating work results. The implementation of a dedicated e-learning environment adopting this approach is presented.

Keywords: Work-process-oriented Curriculum (WPOC), Vocational Education and Training (VET), Learning Arena, Learning by Doing

1. Introduction

Since 1996 a new framework for Vocational Education and Training (VET), called “learning arenas” (in German “Lernfelder”), has been introduced in Germany. In the concept of learning arenas, learning situations in schools are related to the actual work activity in a particular occupation and the work-process-oriented knowledge as the pivotal factor in the design of vocational curricula (Rauner, 2007). In a work-process-oriented curriculum (WPOC), the content matter is no longer based on the structure of the reference disciplines, but is oriented towards the actual work requirements (KMK 1999). The classic distinctions of “theory equals school-based learning” and “practical experience equals work-based learning in companies” or “knowledge equals school-based learning” and “know-how equals work-based learning in companies” are to be removed through the orientation of school-based content to the practical requirements of the vocational and professional work (Fischer and Bauer, 2007).

The learning arena approach has been developed in German context with a long tradition in organizing VET in a “dual” mode between schools and in-company training. However, WPOC shifts the focus from “dual” to a stronger work orientation, which makes it easier to transfer and apply to other educational systems that cannot count on long experience with a dual system. WPOC has gained special attention in many countries as an innovative approach to adapt VET to emerging skill profiles. For example, the Chinese government promoted a curriculum reform in VET through adopting and adapting the ideas of learning arenas (Zhao & Xu, 2008).

Recently some efforts have been made to support learning and work tasks in an educational context that is compatible with a learning arena approach by using digital media (e.g. Howe and Staden, 2015). However, the support is limited in delivering instructional materials to students. The work described in this paper aims at developing a technical approach to support online or blended WPOC systematically and, on this basis, at implementing a dedicated e-learning environment.

2. Characterizing Learning Arenas and Identifying Technical Requirements

1007
In this section we analyze the characteristics of learning arenas and identify technical requirements to develop an e-learning environment.

2.1 Structural Learning Processes

From the perspectives of curricular structure, a learning arena consists of several concrete learning tasks called “learning situations”. The learning situations themselves are set up according to pedagogical principles. Through these learning situations it is possible to simulate typical activities from the particular occupational context from the actual work-processes and to connect such processes with traditional educational content. However, they are not exactly identical with tasks at the company since they have to take the educational mission and goals of the vocational schools into account (KMK 1999). While performing an activity, action-oriented learning takes place. This includes phases of holistic acting (e.g., information-gathering, plan, decision-making, implementation, monitoring, and evaluation) and thus facilitates the required holistic competence development process (Tramm and Krille, 2013). In practice, a learning process is usually triggered by a problem resulting from a work situation. Professional tasks or problems in VET are solved like in the real world of work in sequential and logical steps (activities). The activities make up a complete, multi-dimensional work process that copes with a corporate work order, corporate problem-solving or unknown tasks (Spöttl, 2005).

Technically speaking, in order to guide and scaffold learning in the whole process of a work-process-oriented curriculum systematically, it is required that the whole curriculum can be described in a manner that the e-learning environment can support learners to perform activities one by one following the structural process model.

2.2 Action-oriented Learning

When learning is linked to work-processes, it has to be designed in a way that the competence of coping with today’s challenges at the workplace grows (Spöttl, 2008). Therefore, the core objectives of the newly structured occupational profiles and of the curricula is that learning through experience should develop from informal, unplanned learning into a self-directed and constructive mode of learning (Spöttl and Windelband, 2013). Didactic reference points are situations that for the exercise of the profession are important for learning. These can be summarized by the following points (Niedersächsisches Kultusministerium, 2001):

- Actions to be executed as automatically as possible by the learners are information-gathering, planning, performing, checking, and finally evaluation.
- Actions should be a holistic capture of the professional reality, including technical, safety-technical, economic, legal, ecological, social aspects.
- Actions need to be part of the learners’ integrated experience and need to be reflected in relation to their social impact is reflected.

It is required that the e-learning environment can facilitate learners to conduct actions in real or simulated work situations.

2.3 Acquisition of Knowledge and Skills within Application Context

Learning arenas are thematic units that are oriented towards vocational tasks and procedures. The thematic units can also be included within learning arenas from an academic point of view. Conveying orientational knowledge, system-oriented action and networked thinking are promoted particularly within an action-oriented classroom. It is therefore indispensable that the respective work and business processes are provided with the relevant academic background information (KMK, 1999).

2.4 Outcome-based Assessment

Assessment in vocational learning usually requires the learners to demonstrate that they can carry out the tasks required by their area of study, and that they understand why they are working in this way (Niedersächsisches Kultusministerium, 2001). It is a typical outcomes-based assessment (OBA) that focuses on the results or learning outcomes to be achieved rather than on the content of the learning.
One of the strongest arguments in favor of outcomes based assessment is that it is inclusive, offering clear standards and valid processes through carefully specified outcomes which identify knowledge and practical skills and open the door to conventional vocational assessment practices as in the workplace (Niedersächsisches Kultusministerium, 2001).

3. Technical Approach to Deliver WPOC in a Blended or Online Scenarios

Based on the requirements identified in the last section, we developed an approach to support learning by doing within an online or blended WPOC. The approach is inspired by the ideas of IMS Learning Design (LD) (Koper and Tattersall, 2005), an international e-learning technical standard. Learning design has emerged as a distinct field of research, which is concerned with the development of methods, tools, and resources for helping learning designers in their design process. By adopting an approach of a pedagogy-specific learning design language (Miao, et. al. 2014), we develop a scripting language for representing a WPOC. If a WPOC has been described as a formal process model by using such a scripting language, the computer can configure learning environments and scaffold the learning process. This section will present how our approach meets the requirements identified in the last section.

3.1 Guide and Scaffold Learning Processes Using a Formal Process Model

As mentioned above, we develop a scripting language for modeling a WPOC. The conceptual model of this scripting language is illustrated in Fig. 1. Using this scripting language, a learning arena can be specified by setting values of the attributes of the learning arena (e.g., title, description of the typical professional task, learning objectives, prerequisite, organization of the occupation, objects of the occupation, time schedule, methods, requirements of the occupation, and assessment standard) and by defining a sequence of learning situations, from simple situations to complex situations aligning with cognitive development. A learning situation represents a learning task that was pedagogically transformed from a concrete work task. A learning situation is specified by setting values of attributes of the learning situation (e.g., title, learning situation description, time schedule, learning objectives, learning content, difficulties and important points, and assessment standard) and by defining a set of learning phases, a set of learning activities, or a set of their combinations. From the perspectives of curricular structure, a learning phase is consisted of a set of coordinated learning activities. Whether a learning situation should be represented as phases or as activities depends on the characteristics and granularity of the corresponding learning task. As the elementary unit of a curriculum, a learning activity has to be specified by describing attributes such as title, roles, tool, completion condition, and a concrete task towards an assessable artifact such as writing a report, make a learning plan, generate a solution, make a performance in real- or simulated work environment. In order to fulfill the task, work guidance and information resources about associated theoretical knowledge and practical skills are defined in the activity. After a learning arena is represented as a formal process model using the scripting language, the e-learning environment can guide and scaffold the learning process according to the model by using an enactment service, or called engine.

![Figure 1. The Conceptual Model of a Learning Arena.](image-url)
3.2 Enable Doing by Integrating Application Tools for Work

While executing the process model, the engine enables learners to perform tasks in the authentic or simulated environments. In today’s digitalized world, digital media are deeply integrated into the work processes due to an increased application of information and communication technologies (ICTs). The tools such as Word, PowerPoint, and Excel are usually used in work-processes, so they are integrated within the e-learning environment for supporting the most activities. Meanwhile, certain domain-generic tools such as Concept-map and Mind-map are integrated in the e-learning environment for conducting action-oriented learning. In addition, the e-learning environment provides mechanisms to integrate some third-part domain-specific application tools based on an international standard -- Business Process Management Notation (BPMN) for performing jobs in various fields. For other domain-specific application tools that have no standard interface, our approach suggests to integrate with virtual machines in which those application tools can be deployed and used. Furthermore, in order to support learning through a form-based workflow, the e-learning environment can transfer a form or certain data items within a form from one activity to another.

3.3 Support Learning by Embedding Associated Learning Materials in Application Situations

While performing work-oriented activities, interdisciplinary aspects (e.g. technology, economy, ecology, and law) may be needed. Rather than structuring learning content according to scientific rationality in normal discipline-based curricula, our e-learning environment enables teachers to organize and provide information fragments to represent multi-disciplinary knowledge and various practical skills needed for performing the current activity. Our approach suggests that each activity will be clarified by describing the activity goal and expected artifact such as an analysis or a solution. The instructions and guidance about how to carry out the activity are described in general. If the learner can complete the activity and submit the artifact successfully, the learner is competent to carry out the activity. When the learner lacks of knowledge or skills for solving the problem, she or he can find associated information chunks about theoretical knowledge and practical experience arranged by the teachers in the activity or search by themselves.

3.4 Conduct Outcome-based Assessment by Collecting and Evaluating Work Results

As mentioned above, in each activity certain a work-related application tool should be integrated. The Word, PPT, and Excel tools are used frequently to collect work results such as a plan, an analysis, a solution, a survey, or a work report. Sometimes, professional application tools such as AutoCAD can be used for understanding the job or designing a part. The e-learning environment can collect the work results in the form of files and data as evidences. Various forms of assessment such as self-assessment, peer-assessment, and teacher assessment can be supported by the e-learning environment for judging the competence levels. It is important to note that in blended learning the learner may do performance in physical environments. Assessment can be conducted based on direct observation or the digital documents such as video clips or pictures that records the performance.

4. Implementation of a Web-based E-Learning Environment

By adopting the approach described above, we implemented a web-based e-learning environment for supporting learning by doing in a WPOC. As illustrated in Fig. 2, the core system of the e-learning environment consists of mainly four modules: learning material management, virtual machine management and application tool-set, curriculum authoring tool, and curriculum player. Teachers can upload digitalized learning materials categorized by attributes such as subject-structured topics and media-types to the repository. Some work-related tools are integrated into the system. When a teacher designs a WPOC, she or he can define the curriculum structure as a diagram by using the curriculum authoring tool. When defining an activity the teacher can specify how to fulfill the task as guidance, which application tool will be used, and which artifact is expected to produce. Accesses to information
chunks should be clearly linked to learning materials in the repository. Because of the limitation of the space of this paper, this paper focuses on presenting the curriculum player.

The function of the curriculum player is to deliver an online or blended WPOC. The curriculum player consists of a curriculum engine at server side and a user agent at client side. After a WPOC is specified using the curriculum designer, a set of rules describing the learning process is coded by using a customized Business Process Management Notation (BPMN) (called a curricular scripting language). Interpreting the curricular script, the curriculum engine will guide and scaffold the participants involved in the learning process to remain compliant to the given rules. The user agent will render the user interface according to the information received from the curriculum engine.

In Fig. 3, the user interface (UI) of the curriculum player is shown within a web browser. The tool has three columns: learning space, work space, and activity navigation bar. The activity navigation bar lists executed phases and learning activities of the learning situation. Currently, the user is taking a course “Advanced Numerical Control Machining” and working in the third learning activity that is highlighted in the activity navigation bar. The learning space instructs the learners how to fulfill the current task (“Collect information from a geometric model”). In the work space a text-editor with an initiated table for performing the task is available according to the script. It is expected to fill the table with data collected from a geometric model. In order to support user to work anytime, anywhere, in any machine with a browser and internet connection, a virtual machine with the deployment of an AutoCAD with a template is arranged. If the user clicks the virtual machine button (the icon of a machine) on the left side of activity navigation bar, a virtual machine window will open and present to the user as a pop-up window. The user can work on the virtual machine without the need to prepare work environment, because the work environment has been deployed based on an image of a virtual machine as a template. So, the user can concentrate on analyzing and gathering the data by using the AutoCAD. If the user
doesn’t know how to collect information from the geometric model, she or he can click the links (marked blue) to access information chunks about “how to analyze 3D model of a part” and “how to adjust the orientation of the part”. Then, the associated learning materials will be presented to the user. While performing the activity, the user can discuss with others who are in the same course class or in the same project team by switching between chat-rooms. After the learner submits the fills the table by clicking the button at the bottom, the following activity will be triggered according to the formal process model of the curriculum. It may be an “analyze process arts of the part” or a “peer assessment activity” according to the definition of the script.

5. Conclusions

In this paper, the authors analyzed the characteristics of the learning arena, identified technical requirements, and develop an approach to support learning by doing within a WPOC. The approach can be characterized as: guide the learning process according to the process model representing a work-process-oriented curriculum, enable doing by integrating work-related application tools, support learning by embedding associated learning materials in application context, and conduct outcome-based assessment by collecting and evaluating work results. Furthermore, the implementation of an e-learning environment by adopting this approach was presented. The e-learning environment supports action-oriented learning that is different from a subject and science-orientated system of didactics.

References

Developing Pre-service Teachers’ 21st Century Teaching Competencies via Digital Storytelling

Wan-Lin YANG

Center for Teacher Education & Institute of Education, Cheng Kung University, Taiwan
wlyang9@mail.ncku.edu.tw

Abstract: This qualitative study attempts to explore the potential of digital storytelling for developing pre-service teachers’ 21st century teaching competencies. Digital Storytelling is used as a means to understand pre-service teachers’ professional identity in the current wave of education reform as well as to unpack its potential in the development of pre-service teachers’ abilities of critical reflection and critical technology integration. Findings may provide insights for teacher preparation programs for designing courses that aim at developing pre-service teachers’ 21st century teaching competencies.

Keywords: Digital storytelling, professional identity, critical reflection, teacher education, 21st century teaching competencies

1. Introduction

Current education reform in Taiwan advocates learner-centered learning and digital classrooms. This advocacy leads to the demand for transformation of pedagogical design that supports the development of 21st century skills, such as critical thinking, collaboration, communication, creativity (namely 4Cs), ICT literacy, and life and career skills (Partnership for 21st Century Learning, 2002). Although a considerable number of studies have been made on how to foster students’ 21st century skills, little attention has been given to how teachers can be educated to become qualified 21st century teachers. In response to this paradigm shift and new demands placed upon teachers, teacher education should step further to tackle on what approaches best facilitate teachers to develop 21st century teaching competencies.

2. Objectives of the Study

Using digital storytelling (DST) as a tool, this research attempts to investigate pre-service teachers’ professional identity in the current wave of education reform as well as the potential of DST in the development of pre-service teachers’ abilities of critical reflection and critical technology integration. According to Sachs (2005), teacher’s professional identity stands at the core of the teaching profession. It provides a framework for teachers to construct their own ideas of ‘how to be’, ‘how to act’ and ‘how to understand’ their work and their place in society. While educational reform reshapes what we understand professional knowledge and good practice, it is essential for pre-service teachers to reflect on how they see themselves as a teacher in the 21st century educational current. However, in the field of education, we often encourage students to reflect on their learning experiences without providing tools for reflection. Therefore, one of the research objectives of this study is to use DST as a reflective tool to understand pre-service teachers’ professional identity in the current wave of education reform.

Besides, as emerging technology is introduced and promoted in educational settings, educators nowadays need to have skills and competencies to design, develop, implement and evaluate technology-enhanced curriculum that reflects the needs of 21st century learners. However, research has shown that teacher preparation programs in Taiwan tend to focus on teaching functions of computer software without introducing how technology can be appropriately integrated into instructional design (Chang & Weng, 2006). This often leads to an overemphasis on the introduction of the tool itself rather
than on the potential of the tool to stimulate knowledge construction, meaningful learning, critical thinking, or reflection. Therefore, in this study, pre-service teachers were situated in a technology-enhanced, collaborative learning environment to explore new technologies, to think critically on the affordances of tools, and to reflect on their learning experiences. Another objective of this study is to investigate the potential of the DST curriculum for nurturing pre-service teachers’ deeper understanding of the educational value of DST as well as their competencies of critical technology integration.

In short, the research focus of this study lies in an understanding of pre-service teachers’ perceived professional identity in the current wave of education reform as well as the potential role of DST for the development of their critical reflection, and technology integration competence. Such understanding is pivotal in the search of effective approaches to nurture pre-service teachers’ 21st century teaching competencies as well as serving as a usable basis for designing and developing teacher education curricula to prepare pre-service teachers to become 21st century practitioners.

3. Literature Review

The concept of meaningful reflection has been highly addressed by teacher educators and researchers in the field of teacher education. There has been a growing literature on reflection in the context of teacher education associated with the discussion of portfolios, both the traditional portfolios (e.g. Campbell et al., 2004) and e-portfolios (Wetzel & Strudler, 2006). However, there is a limited body of literature illuminating the use of new technologies to facilitate pre-service teachers’ reflection in the process of their development of their 21st century teaching competencies. Moreover, according to Graham & Phelps (2003), the idea of reflection has been addressed in the field of teacher education, but it is often referred to as reflection on classroom experiences and instructional competencies. It is rarely used as a means for the purpose of self-conscious understanding of oneself as a teacher. Kearney (2011) also argued that learner-centered digital video production often emphasize on technical aspects when developing guideline for educators; however, important educational issues, such as teacher roles, peer learning structures and support for reflective processes, are less the focus.

There are many definitions of digital storytelling. In this study, the researcher refers to the form defined by the Centre for Digital Storytelling in Berkeley, California (Lambert, 2002). This model weaves text, images (e.g. photographs), video, audio recordings of music and narrative together into a 2-6 short multimedia-based artifact with an emotional content to convey personal stories or to transfer knowledge and values. According to Boase (2008), the process of making digital storytelling involves judgment on what and how the story to be presented, that includes a critical selection of or omitting certain information, creating a sequence, and conveying message from a particular angel. This process of comparing, selecting, inferring, arranging and revising information implies the skills of critical thinking and communication that the storyteller needs to have in the creation of the digital storytelling. As in the process of making digital storytelling, one needs to make sense of one’s life experiences before sharing personal narratives, DST can be utilized as an effective approach to access to pre-service teachers’ cognitive process of the development of their teaching profession. According to Tendero (2006), DST allows learners to reconstruct meaning through the process of reflection on experience, therefore, it becomes a powerful tool for self-discovery, self-reflection and for investigating issues of identity.

4. Research Questions

This study attempts to answer a big question: what is the potential role of DST for 21st century teaching competencies? To arrive at the answer to this research question, four sub-questions will also be addressed:

- What are pre-service teachers’ professional identity in the current wave of education reform?
- What are pre-service teachers’ experiences of using DST to explore and to define their professional identity?
• How does the DST project impact pre-service teachers’ ability of critical reflection?
• How does pre-service teachers’ DST experience impact their competence with technology integration in education?

5. Research Site: The Learning Environment, Curriculum and Pedagogy

This research was conducted in a technology integration course in the Center for Teacher Education in Taiwan. The objective of the course is to prepare pre-service teachers for critical technology integration in their future instruction as well as to develop their higher order thinking ability, such as collaboration, leadership, creativity, and the ability of critical thinking, decision making, problem solving and critical reflection. In this course, 6 types of multimedia software (e.g. web development tool, animation software, graphic editing software, video editing software, screen-capture video tool, assessment software), 3 educational software packages (e.g. mind mapping software, a learning management system, an interactive whiteboard), one inquiry-oriented instructional strategy (WebQuest), and one instructional tool (digital storytelling) were introduced by the instructor. Digital storytelling on related educational issues is one of the projects students need to complete throughout the whole semester. Classes had face-to-face meetings for two hours per week for 18 weeks. A Learning Manage System (LMS) was used and served as an extended learning opportunity for interaction, connection, communication, peer feedback, information/resources sharing, and disseminating their finished projects.

A successful technology-enhanced curriculum designer should be able to articulate the interrelationship among affordances, learning theories and pedagogy. To achieve the aforementioned course objectives and to immerse pre-service teachers in a reflective learning context, each educational tool introduced in this course was not taught as separate and independent domain. Instead, the affordances of each tool for teaching and learning was also explored, discussed and modeled, in the hope that such curricular and pedagogical design can help to develop pre-service teachers’ understanding of the complex relationships between pedagogy, content, and ICT. Eventually, they will be comfortable and confident to integrate emerging technologies in meaningful ways to support their future curriculum objectives and pedagogical activities. Throughout the whole training process, pre-service teachers were constantly reminded to be attentive to the questions regarding the affordances of the tools, what they are using, why they are using it, and how they are using it. These questions are crucial to develop 21st century teachers’ competencies of technology integration, and the ability of critical reflection.

6. Research Procedure

In the first week of the course, a reflection paper was assigned for the enrolled pre-service teachers to jot down their thoughts in terms of their teaching philosophy, what constitutes good teaching and their perceptions of technology integration for pedagogical purposes. During the semester, the concept and features of digital storytelling were introduced through demonstrations of different examples of digital stories. Students were asked to identify strengths and limitations of each example, how it can be improved, and how digital storytelling might be used in real classrooms. Pre-service teachers had ample opportunities to explore other educational software and technology-enhanced instructional approaches throughout the whole semester. Toward the end of the semester, they were asked to revisit the questions handed to them in the beginning of the semester again. Then they used DST as a tool to reflect on their changes with respect to the question of “how do I see myself as a teacher in the 21st century educational current?”

PowerDirector 15 was taught to assist pre-service teachers to create their DST project. Pre-service teachers were then asked to write a script, and locate graphics, images, and music they need for the creation of their DST project. Pre-service teachers needed to present their digital storytelling project, to explain the major concept they want to convey, to identify the affordances of the technologies that were involved during the production, to share challenges they encounter, to elaborate
on strategies that they used to solve the problems, and to reflect on what they had learned from this learning experience. Peer evaluation and self-assessment were conducted in a group discussion format.

The DST project aims at unpacking the pedagogical potential of DST by having pre-service teachers answer the question of “how do I see myself as a teacher in the 21st century educational current?” In the process of developing the DST project, pre-service teachers were involved in a series of learning that include reflecting critically on their perceived professional identity, reflecting on their learning experiences; adopting educational tools and instructional approaches to critical implementation of technology for pedagogical purposes; a critique of taken-for-granted assumptions regarding teaching and learning, self-evaluation and peer mentoring. The process of making story from experiences for the purpose of reflection can be a heuristic way to help learners to explain, understand, make sense of and make meaning for their experiences. When the story is shared and disseminate, it can also serve as a means to communicate, interact, and connect with others. Ultimately, by engaging pre-service teachers in the creation of the DST project, they will have the opportunity to develop their 21st century teaching competencies.

7. Participants Selection, Data and Analysis

Participants in this study are secondary pre-service teachers. Participants were selected by way of purposeful sampling (Bogdan & Biklen, 1998). Pre-service teachers whose DST project demonstrates deep reflection and better ICT literacy were chosen to be participated in this study. A multiple case study that involves eight individual case studies was conducted. Each individual participant is a pre-service teacher enrolled in the technology integration methods course, and each participant is considered as one case. A cross-case analysis was presented to inform per-service teachers’ experiences with DST for the development of their professional identity and 21st century teaching competencies.

For the data collection, case study involves multiple sources of information. Therefore, data collected for this study includes a) interviews of the enrolled pre-service teachers; b) document analysis on the reflection paper and digital storytelling projects, which are considered as part of the course requirements, created by per-service teachers; c) participant observation on how those pre-service teachers present and reflect on their digital storytelling project; d) direct observation of the in-class discussion and peer-evaluation on each other’s digital storytelling projects; e) document analysis on the instructor’s teaching log. Such understanding is pivotal in unpacking what learning context or guidance needed to be provided when using digital storytelling as a learning tool to nurture pre-service teachers’ 21st century teaching competencies.

For the analysis of the data, an in-depth description of each individual case (a holistic analysis) is included. Then the researcher focused on the emerging themes (analysis of themes) to have a better understanding of the complexity of the case. This was done within one case (within-case analysis), and then a thematic analysis across the cases was done (cross-case analysis) (Miles & Huberman, 1994).

8. Preliminary Findings and Conclusion

DST in this study was used as a means to access pre-service teachers’ professional identity as well as to unpack its potential role for the development of pre-service teachers’ 21st century teaching competencies. Findings specifically center on pre-service teachers’ perceptions of and experiences with DST in assisting us to understand how DST impacts pre-service teachers’ critical reflection, particularly on the construction of professional identity. In addition, in order to develop effective strategies to nurture pre-service teachers’ 21st century teaching competencies, the researcher delved into pre-service teachers’ perceptions of the DST curriculum introduced to them for designing ICT integrated lessons. Major findings were as follows:

- Pre-service teachers see themselves as facilitators, motivators, companions and inspiration for students in learning.
- Their experience with DST helps them to develop their professional identity.
Developing a digital storytelling project helps pre-service teachers to organize their thoughts and to reinforce their motivation on being a teacher.

Peer evaluation and self-assessment improve the quality of pre-service teachers’ final products.

Creating a digital storytelling project helps pre-service teachers to develop their critical reflection ability.

Pre-service teachers demonstrate great interests in transferring their experiences of the DST project into their future classroom.

This project helps pre-service teachers to develop their confidence on successful technology integration for pedagogical purposes.

Pre-service teachers demonstrate a more critical perception of technology integration. They focus more on the complex relationships between pedagogy, content, and ICT.

The critique and discussion on digital storytelling examples help pre-service teachers critically think about issues of technology integration (e.g. affordances of the tool, integration strategies, pedagogy).

Findings from this study may provide insights for teacher preparation programs for designing and structuring future technology integration courses that aim at developing pre-service teachers’ 21st century teaching competencies.

Acknowledgements

This study was supported by the Ministry of Science and Technology, Taiwan, under the Grant Number of 106-2410-H-006 -058 -MY2.

References


Current Situation of Chinese Primary and Secondary Innovative Teachers’ Evaluation in Maker Classes

Jun-Hao SHAN a* & Shuai-Shuai LI b
a Faculty of Education, East China Normal University, China
b Department of Educational Information Technology, East China Normal University, China
*1191266481@qq.com

Abstract: Maker education is a new educational pattern in China, which also caters for the political ideology called “mass entrepreneurship and innovation”. Maker education pays much more attention to the students’ learning processes, so it is essential for teachers to use proper evaluation to judge the students’ learning situation and behaviors, so as to give students more detailed direction to get better. So investigating innovative teachers’ evaluation competences is an important work to promote the development of maker education. This study tries to analyze teachers’ attitude and progress when evaluating students and find some problems they have faced. A depth interview went behind the questionnaire to find whether teachers were having problems when evaluating students. At the end of the paper, some detailed suggestions are given to help the teachers evaluate students more effectively.

Keywords: Maker education, maker literacy, evaluation situation, teacher’s professional development, evaluation methods

1. Introduction

In China, a new learning pattern called “maker education” has drawn public concern in the round of creative education reform. Maker education draws the advantage of information technology, and becomes a new fashion of creational education (Zhu, 2015). Innovative teachers are a new teacher type in primary and secondary schools, and their goal is to teach students how to use modern open-source hardware and software (such as 3D printer and Scratch) to make creative things. Evaluation is an essential part in the teaching process, however, teachers do not place evaluation in an important position. This study is aimed at assessing the primary and secondary innovative teachers’ evaluation competencies in the process of instruction design to find whether they have some problems when evaluating students and trying to give pointed suggestions to help teachers improve their evaluation skills and build a better teaching environment.

2. Theoretical Frameworks

2.1 Maker Education in China

In China, maker education is a new teaching method. Theoretically, Zhu (2015) stated that maker education was based on the fusion of information technology, inheriting experiencing education, project-based learning and creative education and DIY idea. Maker education is an educational form to foster students’ maker literacies and spirit (Zhu, 2015). In practical perspective, Xie (2016) in Wen Zhou high school has designed a series of maker education curricula to teach students how to use open source hardware to make creative products. Wu (2016) has used SCRATCH in the class to teach students how to code on the computer.
2.2 Maker Literacy in Maker Education

Maker literacy is in the setting of actively responding to “mass entrepreneurship and innovation” from Ke-Qiang Li, the Premier of State Council. Zhu (2015) stated that maker literacies consist of five dimensions: the ability of finding, analyzing and solving problems, making creative production, social competence, thinking ability and creative passion. Wang (2017) gave a clear outline about maker literacies which are also classified into 4 major dimensions. Through summing up and analysis, each of the literacy corresponds to one particular instructional objective. The following graph shows the correspondence of maker literacies and instructional objective. By knowing what students need to achieve in innovative classes, teachers can know better what to teach and how to evaluate students properly.

![Figure 1. Map of Maker Literacies.](image)

2.3 Evaluation Patterns: Summative and Formative

Evaluation is indispensable of instructional process. Teaching evaluation includes not only the summative evaluation, which is represented by a standardized test, but also a formative evaluation that aims at the learning and focuses on learning process (Leung & Mohan, 2004). The summative evaluation is an important means of testing the teaching results, but it cannot evaluate the teaching process, which is a critical part. However, the formative evaluation makes up for these weaknesses. The functional comparison between them is described in figure 1.

![Figure 2. Map of Evaluation Functions (Murphy & Torrance, 1990).](image)

2.3.1 Formative Evaluation

Formative evaluation is a pre-planned "process", and teachers in this teaching process continue to pay attention to and monitor the student's learning situation to collect evidence of student learning to amend the teaching (Stiggins, 2005).
The essential feature of formative evaluation is feedback and improvement (Zhao Decheng, 2013). Mo Yan (2004) thinks formative evaluation stresses the following features: the purpose is formative; the subject is extensive; the content is comprehensive; the means are flexible and diverse; the results are qualitative and quantitative.

From figure 1, it can be seen that formative assessment can offer effective diagnosis of problems in educational practice and effective feedback as well as guide educational activities to improve teaching quality. So, formative evaluation is a key and an important method of high quality teaching.

2.3.2 Summative Evaluation

Summative evaluation refers to the cumulative evaluations, usually occurring at the end of a unit or topic coverage, which intends to capture what a student has learned, or the quality of the learning, and judge performance against some standards.

The purpose of summative evaluation is various. It is to pass or fail a student; to grade or rank a student; to allow progress to further study; to assure suitability for work; to predict success in future study and work etc.

3. Research Methods and Data Analysis

3.1 Research Methods

In the case when assessing teachers’ evaluation and implementation, competences seem to be either objective or subjective. No one can describe teachers’ evaluative process better than themselves, so this study mainly focuses on investigating and interviewing teachers’ subjective feeling about their evaluation competences situation. So questionnaire survey and interview are the main research methods. How to interview teachers’ evaluation and implementation competences roundly and authoritatively is a crucial problem to solve. IBSTPI published the latest authoritative instructional designer competences scale for teachers in 2013, and evaluation and implementation is the last part of the whole scale. This part contains 3 dimensions and 14 detailed standards. It is no doubt good criteria for researchers (the author) to interview and make a questionnaire to find the unsolved questions. The step we followed, which were adapted from (Koszalka, 2013), are detailed below.

(1) **Creating the Database.** Data from this study were mission statements from innovative teachers who had some teaching experiences in primary and secondary schools in China (n=19). And the study also chose four teachers who had a wealth of teaching experience for a depth interview. Maker education is still in the initial developing period in China, so there are not abundant expert teachers for us to interview.

(2) **Defining the unit of analysis.** Researchers referred to the evaluation scale from IBSTPI and converted into an interviewing questionnaire. The questionnaire contains the following dimensions: teachers’ perspective about evaluating instructional and non-instructional interventions, revise instructional and non-instructional solutions based on data, implement, disseminate, and diffuse instructional and non-instructional interventions. And then in consideration of a better understand of the meaning of each questions, researchers annotated the specialized vocabularies to be got across accurately.

(3) **Developing categories and a coding scheme.** Research team members chose Likert scale to measure the teachers’ subjective assessment towards their evaluation competences in a quantitative way. And there were also some open-ended questions for the interview after the questionnaire to see whether they had run into a stone wall when evaluating the students.

(4) **Data analysis and conclusion drawing.** The results of data analysis were was presented in two aspects: quantitative description and qualitative description.
3.2 Data Collection and Analysis

The sample is mainly from the 19 front-line teachers who come from Shanghai and Beijing, including 16 female teachers, 3 male teachers. In this study, questionnaires and in-depth interviews were used to investigate. And the effective rate of questionnaire and in-depth was 100%. The questionnaire is labeled with the 5 points Likert Scale, and the options are "fully compliant, more consistent, generally consistent, less compliant, totally incompatible" with a score of 5, 4, 3, 2, 1, and analyzes the results through SPSS 23.0.

Table 1: Evaluate Instructional and Non-Instructional Interventions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.90</td>
<td>2.79</td>
<td>2.26</td>
<td>2.84</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.87</td>
<td>.78</td>
<td>.65</td>
<td>.83</td>
</tr>
<tr>
<td>Range</td>
<td>3.00</td>
<td>2.00</td>
<td>3.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Total Average</td>
<td></td>
<td></td>
<td></td>
<td>2.70</td>
</tr>
</tbody>
</table>

The above four indexes investigated whether the innovative teachers set up evaluation intervention in the teaching process. The survey finds that most of the innovative teachers will take the initiative to design a set of evaluation programs and publish evaluation reports (total average = 2.6974). This shows that most of the innovative teachers have the awareness and ability of evaluation. In the choice of evaluation methods, more innovative teachers incline to use the formation of evaluation.

Table 2: Revise Instructional and Non-Instructional Solutions Based on Data.

<table>
<thead>
<tr>
<th>Index</th>
<th>Identify Product And Program Revisions Based On Review Of Evaluation Data</th>
<th>Revise The Delivery Process Based On Evaluation</th>
<th>Revise Products And Programs Based On Evaluation Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.21</td>
<td>2.05</td>
<td>2.16</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.42</td>
<td>.78</td>
<td>.69</td>
</tr>
<tr>
<td>Range</td>
<td>1.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Total Average</td>
<td></td>
<td></td>
<td>2.14</td>
</tr>
</tbody>
</table>

These three indexes investigated the ability of teachers to reflect students’ internal feedback. The study found that the overall mean of the dimension was low (total average = 2.1403). This shows that most innovative teachers do not form a good sense of reflection, and cannot do it according to the evaluation of teaching feedback to amend their own evaluation program, and do not really play the role of the implementation of evaluation programs.

Table 3: Implement, Disseminate, and Diffuse Instructional and Non-Instructional Interventions.

<table>
<thead>
<tr>
<th>Index</th>
<th>Change Goals</th>
<th>Plan for Diffusion</th>
<th>Disseminate Interventions</th>
<th>Monitor Implementation</th>
<th>Identify Required Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.31</td>
<td>2.68</td>
<td>2.74</td>
<td>2.74</td>
<td>2.47</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.95</td>
<td>.75</td>
<td>.61</td>
<td>.56</td>
<td>.84</td>
</tr>
<tr>
<td>Range</td>
<td>4.00</td>
<td>3.00</td>
<td>2.00</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Total Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.45</td>
</tr>
</tbody>
</table>
These indexes reflected the ability of innovative teachers to implement, promote and reflect on external feedbacks. We can see from the third and fourth indexes that the teachers now have a certain awareness of the implementation and promotion (mean > 2.50). However, there are significant differences in the ability of founding teachers to integrate knowledge goals and skills goals into the ultimate goal of teaching (standard deviation > 0.5).

And the second, third and seventh indexes show that most of the teachers did not have strong external reflection awareness and ability after the evaluation and promotion, so that they can not get timely and effective development of targeted strategies to adjust the evaluation program (mean ≈ 2.20).

3.3 Conclusions

The overall average of the three dimensions is shown above. It can be seen that the overall average of the three dimensions is between 2.0 and 3.0. Corresponding to the Likert scale, it indicates that the overall evaluation of the majority of Chinese innovative teachers is at a middle level now. From the figure 3 we can draw conclusions as follows:

1. There are individual differences among teachers in evaluation process (standard deviation >0.5 for almost all indexes).
2. Most teachers can plan and use evaluation in maker education.
3. Most teachers can implement and diffuse their evaluation methods.

Teachers lack having reflection to the evaluation results and can not make good use of it

Figure 3. Map of Teacher Overall Evaluation Situation.

4. Suggestions

After combining the depth interview with the data analysis results (due to space limitation, depth interview has been already blended into the suggestions), the research team raised the following four suggestions for teachers to refer.

4.1 Establish Evaluative Scales for Students’ Multiple Competences

In order to avoid evaluating students subjectively, some standard evaluative scales need to be established for teachers to find the students merits and shortcomings objectively. There are already some existing authoritative evaluative methods to measure students’ competences. Take creativity as an example, creativity is complex, which has many facets and occurs in all domains of life. Lucas (2016) raised five-dimensional model of creativity which contains inquisitive, imaginative, persistent, collaborative and disciplined and its detailed evaluative indexes. Meanwhile, each school has its own philosophy of education, and schools can organize innovative teachers to formulate a characteristic evaluation scales to demonstrate philosophy of schools.
4.2 Make the Evaluation Work Diversified

Evaluation diversification can be divided into two dimensions: valuator and evaluative activities. Firstly, teachers cannot evaluate students’ competences thoroughly. So, try to combine teacher evaluation with self-evaluation and peer-evaluation. Self-evaluation can let the teacher know what the students really think about themselves and peer-evaluation can give students various improvement suggestions from other students. It is also a good chance for students to share their ideas and broaden their horizons. Secondly, using evaluative scales for students merely seems to be too boring and make students lose interest in the innovative classes. There are many activities to make evaluation more effectively such as class presentation and holding science and technology festivals in schools. Class presentation can give the students opportunities to show their ideas and products. It can also give students a platform to talk about the personal evaluation about their ideas, and detailed advice from both teachers and other students can also be given to the exhibitors.

![Figure 4. Different evaluation pattern.](image)

4.3 Use Portfolio Assessment to Measure Students’ Competence Enhancement

Portfolios assessment collects more than a diverse body of finished work. In fact, they gather what we have come to call biographies of works, a range of works and reflections. When students are asked to return to their collections of work, finding what has changed with time and what still remains to be refined. This learning process can help students realize their progress. It is a good evaluation method to record the phased achievements in students’ personal learning process. It also helps teachers find students merits and shortcomings from different aspects. Innovative teachers can give each student a portfolio to allow them to put every idea and report in it. To be frankly, it is also a memorial thing for students and their parents.

References

Students’ In-Class Answering Activities on Facebook: Effects on Participation, Learning Satisfaction and Anxiety

Yu-Hsin LI & Fu-Yun YU
*Department of Civil Engineering, National Chi Nan University, Taiwan
Institute of Education, National Cheng Kung University, Taiwan
fuyun.ncku@gmail.com

Abstract: The effects of students responding to the teacher’s in-class questioning on Facebook on participation, learning satisfaction and learning anxiety were examined. Based on its integration in one undergraduate engineering course for a whole semester, several findings were obtained. First, more than 85% of the participants felt that their learning attitudes, habits and behavior changed at the ‘tremendous’ and ‘a great deal’ levels after exposure. Second, ‘inviting and equal participation for all’ emerged as the one salient theme the designed activities created, leading to positive effects on various aspects of learning. Third, the majority of the participants associated positive emotional feelings (e.g., enjoyment, satisfaction, accomplishment) and did not associate negative feelings (e.g., worry, fear, pressure, nervousness) with the activities. Based on the results of this work, and in view of the need for active and reflective learners in today’s society, and the low technological threshold and prevalent use of Facebook by undergraduates, suggestions for instruction are provided.

Keywords: Facebook, in-class answering activities, learning anxiety, learning satisfaction, participation, teacher questioning

1. Introduction

Learners are generally accustomed to holding a passive, receiving learning mode while attending lectures, which is reported to lead to inert knowledge (Renkl, Mandl, & Gruber, 1996). As the 21st century skills deemed core capacities to succeed in colleges, careers and life in today’s knowledge-based economic society are better cultivated through active participation and constructivist learning experiences (Pearlman, 2009), ways to change current learning habits warrant serious consideration.

From a constructivist perspectives, the goal of instruction is to provide opportunities and spaces for students to reflect and create their own interpretations of the information received for higher-level cognitive development (von Glasersfeld, 1987). While the potential of teacher questioning for promoting active learning and reflective thinking in classrooms is promising (King, 1995), its effects on ensuring equal participation remain underexplored. In view of this, and the prevalent use of Facebook among college students, issues regarding how Facebook may assist in this context are considered in this study, along with the effects on learning attitudes, habits and behavior. In addition, as students have been reported to show signs of hostility and resistance in response to instructional innovations that require them to take on more responsibilities (Silver 1994), the effects of the designed activities on two affective aspects suggested to have a decisive impact on successful learning—learning satisfaction and anxiety (Krashen, 1987), were examined.

2. Method

Forty-two undergraduates enrolling in a 3-credit hour engineering course participated for a whole semester. A Facebook group was created for the exclusive use of the participants via any personal mobile device. Questions related to the instructional content (a total of 89 question items) were posed in class intermittently to: solicit students’ prior knowledge or past related experience on the focal

1024
topic; assess students’ understanding of the content presented; provide opportunities for students to apply the concepts and principles introduced; and build relevance regarding the learned material and their lives. At the last instructional session, the participants were asked to write a response to one question—‘After exposure to the teacher’s in-class questioning on Facebook, the extent of change you felt/observed regarding your own attitudes toward attending lectures, learning habits and behavior was: tremendous, a great deal, not much, almost none, none. Please elaborate on your selection.’ In addition, learning satisfaction and anxiety scales were completed individually by the students.

3. Results and Discussion

Thirty-eight students completed the end-of-semester survey (90.48% response rate). All respondents felt/observed some degree of change in their own attitudes toward attending lectures, learning habits and behavior, after exposure to the activities. More specifically, the students’ selections ranged from ‘tremendous’ (10.53%), ‘a great deal’ (76.32%), ‘not much’ (10.53%) and ‘almost none’ (2.63%). Not a single participant marked the ‘none’ option. Constant comparison method was applied to the students’ descriptive explanations regarding their felt/observed change, and the results highlighted one salient theme—‘inviting and equal participation for all’ as created by the designed activity. Among the 38 responses received, 22 participants indicated that ‘answering the teacher’s in-class questioning on Facebook’ was less intimidating, and this method provided more time and space for all students to participate in such a way that they were more willing to express and share their views in class with their classmates and the instructor. As a consequence, ‘becoming more participatory in class,’ ‘being more attentive to class,’ ‘reflecting more on materials presented,’ ‘thinking deeper into issues discussed,’ and ‘learning more’ were felt/observed by many respondents.

With regard to learning satisfaction and anxiety, the respondents generally rated their experience favorably for each of the statements on both scales (Table 1). Overall, 60.53~100% and 57.89~92.11% of them ‘agreed’ and ‘strongly agreed’ with the statements on the learning satisfaction and anxiety scales, respectively. Analyses of the summed-up data and each of the statements, using one-group t-tests with 3 as the expected mean, showed that all results were statistically significant.

Table 1: Descriptive statistics (f, %, mean) and t-value of learning satisfaction and anxiety scales

<table>
<thead>
<tr>
<th>Learning satisfaction</th>
<th>1 * 2 3 4 5</th>
<th>4+5 (%)</th>
<th>Mean</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Being able to participate in ‘in-class answering activities on Facebook’ in this course is enjoyable.</td>
<td>0 0 1 15 22</td>
<td>37 (97.37)</td>
<td>4.55</td>
<td>17.47</td>
</tr>
<tr>
<td>2. I like the way Facebook was used for in-class answering activities in this course.</td>
<td>0 0 0 16 22</td>
<td>38 (100)</td>
<td>4.58</td>
<td>19.71</td>
</tr>
<tr>
<td>3. I dislike using Facebook as a platform for in-class answering activities in this course.</td>
<td>1 0 7 19 11</td>
<td>30 (78.95)</td>
<td>4.03</td>
<td>7.51</td>
</tr>
<tr>
<td>4. ‘In-class answering activities on Facebook’ increases my chances of interacting with the instructional materials.</td>
<td>0 0 2 21 15</td>
<td>36 (94.74)</td>
<td>4.34</td>
<td>14.39</td>
</tr>
<tr>
<td>5. Learning through ‘in-class answering activities on Facebook’ is interesting and novel.</td>
<td>0 0 0 14 24</td>
<td>38 (100)</td>
<td>4.63</td>
<td>20.84</td>
</tr>
<tr>
<td>6. I hope other courses can adopt this type of teaching approach for learning.</td>
<td>0 1 6 16 15</td>
<td>31 (81.58)</td>
<td>4.18</td>
<td>9.24</td>
</tr>
<tr>
<td>7. I feel participating in ‘in-class answering activities on Facebook’ is easy and convenient for information-sharing and interaction in this course.</td>
<td>0 1 0 18 19</td>
<td>37 (97.37)</td>
<td>4.45</td>
<td>14.01</td>
</tr>
<tr>
<td>8. Learning via ‘in-class answering activities on Facebook’ satisfies my needs.</td>
<td>0 1 3 23 11</td>
<td>34 (89.47)</td>
<td>4.16</td>
<td>10.65</td>
</tr>
<tr>
<td>9. I am satisfied with my performance in ‘in-class answering activities on Facebook.’</td>
<td>1 1 5 19 12</td>
<td>31 (81.58)</td>
<td>4.05</td>
<td>7.31</td>
</tr>
<tr>
<td>10. Participating in ‘in-class answering activities on Facebook’ gives me a sense of achievement.</td>
<td>0 2 13 15 8</td>
<td>23 (60.53)</td>
<td>3.76</td>
<td>5.59</td>
</tr>
</tbody>
</table>

Learning anxiety
| 1. I felt nervous when I could not answer questions during 'in-class answering activities on Facebook.' | 2 5 9 16 6 | 22 (57.89) | 3.5 | 2.88 |
| 2. I was comfortable and did not feel pressure participating in 'in-class answering activities on Facebook' in this course. | 1 0 6 15 16 | 31 (81.58) | 4.18 | 8.25 |
| 3. ‘In-class answering activities on Facebook’ did not worry me at all. | 0 2 8 16 12 | 28 (73.68) | 4 | 7.18 |
| 4. Participating in 'in-class answering activities on Facebook' did not frighten me | 0 1 4 20 13 | 33 (86.84) | 4.18 | 10.13 |
| 5. It was very stressful to learn through 'in-class answering activities on Facebook' as used in this course. | 2 2 2 23 9 | 32 (84.21) | 3.92 | 5.77 |
| 6. Learning through ‘in-class answering activities on Facebook’ gave me a sense of dread when I attended classes. | 1 1 2 19 15 | 34 (89.47) | 4.21 | 8.64 |
| 7. I often had an uneasy feeling when learning through ‘in-class answering activities on Facebook.' | 0 2 3 20 13 | 33 (86.84) | 4.16 | 9.16 |
| 8. It saddened me when I found out that I would participate in ‘in-class answering activities on Facebook’ in the future. | 0 2 3 14 19 | 33 (86.84) | 4.32 | 9.76 |
| 9. I felt confident and was at ease learning through ‘in-class answering activities on Facebook.’ | 1 0 3 26 8 | 34 (89.47) | 4.05 | 8.96 |
| 10. I was not nervous participating in ‘in-class answering activities on Facebook.’ | 0 1 2 22 13 | 35 (92.11) | 4.24 | 11.44 |

* Negative statement. Scoring on the negative statements was reversed, with higher scores reflecting more satisfied and less anxious.

1: Strongly disagree; 2: Disagree; 3: No opinion; 4: Agree; 5: Strongly agree

In sum, the facilitating effects of the designed activities on learning were well supported in this study. These included: positively changing students’ attitudes, learning habits and behavior; promoting equal participation among the participants; and leading to constructive learning. Moreover, the participants associated positive emotional feelings (e.g., enjoyment, satisfaction, accomplishment) and did not associate negative feelings (e.g., worry, fear, pressure, nervousness) with the activities. Based on the results of this work, and in view of the need for active and reflective learners in today’s society, and the low technological threshold and prevalent use of Facebook by undergraduates, university instructors are encouraged to consider adopting Facebook to support in-class questioning for active, deep, interactive, and constructive learning in their entire class.

Acknowledgements

This study is supported by the Ministry of Science and Technology in Taiwan under contract number: MOST 106-2511-S-260-002-.

References


Student-Generated Feedback for Online Student-Generated Multiple-Choice Questions: Effects on Question-Generation Performance and Perspective-Taking Development

Fu-Yun YU* & Wan-Shan WU

Institute of Education, National Cheng Kung University, Taiwan

*fuyun.ncku@gmail.com

Abstract: An experimental study examining the learning effects of student-generated feedback for online student-generated multiple-choice questions was conducted. Four seventh-grade classes (n=109) participated for eight weeks. An online learning system was extended to support the activities. The results from ANCOVA confirmed additional benefits gained from having students generate feedback for each of the options of student-generated multiple-choice questions for the promotion of question-generation performance and perspective-taking development, as compared to student-generated multiple-choice questions alone. The results were supported and explained in light of the self-explanation and empathy theories.

Keywords: Online learning activity, perspective-taking development, question-generation performance, student-generated feedback, student-generated questions

1. Introduction

The value of student-generated questions for promoting cognitive, affective and social development has been well recognized in a wide range of domains and disciplines (Rosenshine, Meister & Chapman, 1996; Yu & Su, 2015). Moreover, positive relationships between question-generation performance and academic achievement have been confirmed (Yu & Wu, 2012). In view of these findings, a growing number researchers and practitioners are expressing interest in leveraging various pedagogical arrangements for student-generated questions, in order to both increase the quality of student-generated questions and take advantage of the related learning effects.

The type of questions used in this context has been shown to affect learning. In particular, in Yu and Li’s study (2011) the students who engaged in multiple-choice question-generation activities were found to have significantly better performance in higher cognitive-level achievement assessments and better use of the elaboration strategy than those engaged in short-answer question-generation activities. With this in mind, and since the current literature related to mainly involves students generating questions, this work focused on the effects of having students further generate feedback for each of the options of the generated multiple-choice questions that can be used during drill-and-practice (D&P) sessions.

Multiple-choice questions come with a specific structure that may provide added learning benefits. More explicitly, with multiple-choice question-generation the students are given opportunities to generate question stems and three to five alternative answers (consisting of the correct answer and plausible distractors) in accordance with the study material. Guidelines frequently emphasized to students during such a task include generating questions that assess the main ideas of the study material, and providing options that can detect misconceptions and differentiate those learning from those not. On the other hand, providing informative feedback following students’ answers to multiple-choice questions (i.e., explanations provided for the incorrect responses and justifications given for the correct answers), factors related to instructional contexts, students (e.g., prior knowledge and skills, academic motivation) as well as the contents, presentation and functions of feedback, all need to be considered (Narciss & Huth, 2004). As such, it is anticipated that given opportunities to further generate feedback for each of the options of multiple-choice questions, students would be more likely to be prompted to
re-assess their generated question stems along with alternatives and refine the content accordingly to ensure their quality and appropriateness, with regard to their peers’ knowledge and affective states in mind (e.g., level of difficulty and discrimination, wording, use of term, and so on.). In sum, the researchers speculated that through such externalization and verbalization processes prompted by student-generated feedback for student-generated multiple-choice questions, this process should have positive effects on question-generation performance and perspective-taking ability according to Chi, Bassok, Lewis, Reiman and Glaser’s (1989) self-explanation and Davis’s (1983) empathy theories.

2. Method

A pretest-posttest quasi-experimental research design method was adopted. Four seventh-grade classes ($n=109$) participated and were randomly assigned to the two treatment groups—student-generated feedback for online student-generated multiple-choice questions (the experimental group: SGQ+SGF, $n=55$), and online student-generated multiple-choice questions only (the contrast group: SGQ-only, $n=54$). The online learning activities were introduced to support Chinese learning each week in the school’s computer lab for eight weeks. An existing online system (QuARKS) was extended and adopted for the activities.Basically, for SGQ, students generate a question stem, four options (including the correct answer), and an annotation explaining the main ideas being tested. For SGQ+SGF, students provide a justification for the correct answer and explanation for incorrect responses to accompany each of the options of the focal question, which are used as feedback in response to users’ answer during online D&P activities. In QuARKS, multi-media files can be used as content of questions and feedback. For the design principles of the expanded system, refer to Yu and Liu (2016).

This study consisted of four main stages: training, baseline establishment and pretests, experimental intervention, and posttests. During the training stage (one session), the purposes and fundamentals of SGQ and operational procedures of QuARKS were introduced before individual hands-on practice with the system. To establish a baseline of question-generation performance and perspective-taking ability, all participants engaged in the SGQ-only condition during the 2nd stage (one session); therefore, data on the observed variables could be collected and used as covariates. Afterwards, during the experimental intervention stage (five sessions), the participants assigned to different treatment groups were engaged in SGQ+SGF and SGQ-only for the experimental and contrast groups, respectively. A brief training session on feedback-generation was provided for the SGQ+SGF group prior to their first encounter with SGF. During both the 2nd and 3rd stages, quick reviews of the Chinese instruction covered in the current week and whole-class feedback on student performance on the previous activity was given before the participants were directed to individually generate three multiple-choice questions, either with SGF (the experimental group) or without SGF (the contrast group), according to the learning material covered in the current week in QuARKS (35 minutes). At the conclusion of the 8-week study, the participants completed the same perspective-taking scale, and their question-generation performance at the last online activity was used as posttests. Question-generation performance was graded along five dimensions included by most researchers: fluency, importance, flexibility, elaboration, and originality, and inter-rater consistency was ensured ($r = .82$). The cognitive empathy subscale of Davis’s interpersonal reactivity index (1983) was adapted to fit the situation (i.e., SQG) for perspective-taking ability assessment (6-item, Cronbach $z = .89$). Sample questions included: During SQG, I could view the current learning topic and SQG task from other people’s perspective; During SQG, I would ask myself—if I were him/her, how would I feel or think when I review the generated question and options?

3. Results and Conclusions

The descriptive statistics of the two groups on the observed variables were listed in Table 1. The results of the analysis of covariance (ANCOVA) found significant differences in question-generation, $F (1, 106) = 8.01, p = <.05$, and perspective-taking performance, $F (1, 106) = 5.589, p = <.05$, with the students in the SGQ+SGF group having higher scores than those in the SGQ-only group.
When given opportunities to generate feedback for the options of multiple-choice questions, in addition to providing verification of response accuracy or inaccuracy, students provide justification for the correct answer and explanations for incorrect answers. To do so, students are likely to contemplate questions such as: What kind of misconception or incomplete knowledge would the test-taker hold if he/she chose any of the inaccurate options? What kind of information can be provided to rectify such misconceptions? What and how to teach the test-taker to recognize what is presented in the question stem or alternatives that indicate the option is incorrect? On the other hand, for those choosing the correct answer, questions like the following may be considered during feedback-generation process: Can any additional pieces of information be provided to enhance or extend the test-taker’s current learning status? If so, in what form? In view of self-explanation theory (Chi et al., 1989) and empathy theory (Davis, 1983), these thought processes may prompt the question-author to re-examine the questions they have already generated while keeping their peers’ situations in mind if any revisions are made. As confirmed in this study, SGQ+SGF, as compared to SGQ alone, led to better question-generation performance and perspective-taking development. Overall, this study is the first work to empirically examine and yield data confirming the beneficial effects associated with SGF with SQG.

Acknowledgements

This study is supported by the Ministry of Science and Technology in Taiwan under contract number: MOST 105-2511-S-006-005-MY3.

References

Training System for Puncture Operation
Force Adjustment in Hemodialysis

Ren KANEHIRA a*, Kazunori YAMAZAKI a & Hideo FUJIMOTO b
a School of Health Sciences, FUJITA Health University, Japan
b Nagoya Institute of Technology, Japan
*kanehira@fujita-hu.ac.jp

Abstract: The study deals with a training system for the force adjustment in puncture operation for university students in clinical engineering. It is known in hemodialysis treatment that high level of accuracy is required including the force, angle and needle tracks when inserting a needle into the blood vessel, while it is difficult to master the complex technique within a limited university time with fewer exercises. For the reason, we construct a training system on PC having the ability of insertion force response, quantitative operation effect evaluation and good repeatability. The students experience operation in such a vivid environment with real-timely evaluation. The effectiveness of system was proved with questionnaire on students.

Keywords: Computer Training System, Skill Science, Clinical Engineer, Support to operations

1. Introduction

With the progress in computer science and technology, computer-supported education/training systems have obtained increasing attention (Watanabe, et al., 2010). Similarly in clinical engineering, the development of training systems with the most advanced computer technology has been taken into account (Sueda, 2010). A clinical engineer faces a wide range of tasks including operations of multiple medical machines, and the management and maintenance of them. To fulfill the tasks requires complex techniques and rich expertise. However, it is very difficult to get such a large sum of knowledge, skill and experience within the limited university time in which students are always with less operation practice and particularly for frequent trouble-shooting.

In order to solve the above problems, we have carried out a series of studies for construction of teaching/training systems with highly simulated medical operations (Kanehira, et al., 2010). For obtaining higher training efficiency towards the final goal of such a training system with low cost, rich experiences and repeatability, we proposed in this study one for clinical engineers using the up-to-now knowledge from our sequential researches (Kanehira, et al., 2014). In addition to conventional e-learning for general knowledge, special attention was paid to the practical operation on medical machines (Kanehira, et al., 2016).

The conventional e-learning focuses mainly on theoretical knowledge like that in textbook, with less attention paid to how to operate the practical medical machines. The knowledge from textbook are not as good as those obtained from rich experience (empirical knowledge), body movements (embodies knowledge), and tacit knowledge (Suwa, 2009). Therefore, we are focusing the research on the operation training with the physicality in the field of skill science.

We chose a training system for puncture operating force in hemodialysis. It is considered that such a system must be with operating force response capability in hemodialysis. The system was constructed on a VR (virtual reality) environment in such a way that puncture operation drills with force feedback senses can be repeatedly achieved without the use of real machine or on a real human body. In more detail, an arm model with clinical puncture needle capable of force sensitivity was constructed using the Phantom force-feedback device. Operation exercise can be repeated in the VR training environment.
2. Hemodialysis and Dialysis Puncture Technique

Hemodialysis is the operation to take out the dirty blood from a patient with disabled kidney, to purify it through a dialyzer and to put the clean blood back to his body. Each operation normally takes four hours, and the operation must be repeated every 2 days. The operation starts with a “puncture” operation, by holding a needle and stinging into the blood vessel. The puncture operation must be repeated two times for each treatment, including one needle sting to take out the blood from vessel and another to return it to vessel by putting needles into the special shunts put earlier in the blood vessel. The needle should be stung precisely into the shunts at a proper angle under a difficult condition that blood vessels are structurally complex and visually hidden underneath the skin. Therefore, a puncture operation is always more difficult and dangerous compared with normal injections. Furthermore, an erroneous puncture operation may result in accidents such as heavy bleeding. The puncture is one of the most important techniques for hemodialysis.

Undoubtedly, punctures with high levels of precision are required for such operation. However, it is impossible to allow students to do practical puncture on patients during clinical training. As the result, it can be imagined how anxious and uneasy the student may be when he faces a patient for the first time. However, there has not yet been such a training system up to the present time upon our best knowledge. In order to solve the above problems, we proposed the following training system for “puncture” operation for hemodialysis.

3. Puncture Technique Training System

We developed an early version of the system for teaching of insertion angle and track, which was not satisfactory as it did not provide force feedback during insertion. Therefore, we pay attention mainly on the force feedback ability and repeatability of the system which represents the operation sense or feeling of the operator and the training efficiency. In more detail, the training system was constructed as a Virtual Reality (VR) environment on PC. The device “PHANTOm Omni” (Sensible Technology Inc./OpenHapticsToolkit) was used to provide force feedback. A VR arm model was formed, and the reaction force from the experienced clinical engineers was input as teaching standard. The improved system thus possesses the ability of operation sense, insertion angle and direction, and so on. The system was used in training our students, and the effect was confirmed with questionnaire.

For an easy understanding of the insertion angle and distance in VR most like that in real operation, a simple arm model with skin, inner vessel and outer vessel of three layers was constructed, each of which was with force resistance. The size and force resistance of vessels were made changeable in response to the gender, age of difference people. It is required to provide teaching standard data in the system. For this purpose, operations from several experienced clinical engineers were recorded, analyzed and quantified as training standard index. The force resistance of 0.3N, 0.2-0.6N, and 0.4Nf was set for the skin force resistance, inner vessel and out vessel, respectively.

The Ubuntu suitable for ROS (Robot Operation System) was used as the OS for Phantom. The configuration and flowchart for operation and information were shown in Fig.1. When using this system, the students are requested to view first the electronic text of the flow chat of how to do the
puncture operation. Then, they do puncture exercise on PC using Phantom device. The Phantom provide force feedback by force resistance, and the insertion angle was displayed by numerical value and graph. Whether or not the insertion reach the upper layer will be indicated by words and color so the students can confirm their operation easily (Fig.2). They do repeated exercise to improve their skills while confirming the distance, angle, force, etc. during the exercise.

We evaluated the system with 16 participants at our university, including one teacher, 11 male students and 4 female students. A questionnaire was done upon them after finishing the use of our training system. Such questions as “have you experienced puncture with the system”, “do you wish to use the system before going to clinical”, were set and a 5-stage evaluation was asked for their answers. The teacher tell that he did experience the real puncture sense with the training system, and 90% of the students gave positive evaluation answers.

![Figure 2. Experience on puncture operation sense of touch](image)

4. Conclusion

This study deals with the teaching/training of the force adjustment in puncture operation for the clinical engineering university students. It is known in hemodialysis treatment when inserting a needle into the blood vessel of a patient, high level of accuracy is required including the force, angle and needle tracks. For such the reason, we construct a training system on PC with force feedback device, quantitative evaluation and repeatability. The students experience the operation in such a vivid environment, evaluated from the PC and correct their mistakes real-timely. The effectiveness of training was proved with the answers from the trained students.

Acknowledgements

This study was supported by JSPE KAKENNHI Grant Number 17K01441.

References


Mobile Learning in Higher Education in Sudan

Adam TAIRAB a,b, Ronghuai HUANG a,b*, Mohannad TAHA a & Kirk PERRIS d
a School of Educational Technology, Beijing Normal University, China
b Smart Learning Institute, Beijing Normal University, China
c Deanship of Quality and Development, Sudan University of Science & Technology, Sudan
d 1701-25 St. Mary Street Toronto, Ontario Canada M4Y 1R2
*huangrh@bnu.edu.cn

Abstract: This paper examines implementation of mobile learning in higher education in Sudan and whether universities have planned or planning to apply this technology. The paper also aims to discover the factors that lead to impeding the application of mobile learning. The findings demonstrate absence of strategic plan for mobile learning in higher education despite existence of ICT policy in education in the region. Furthermore, the results show some factors hindering use of m-learning as instance lack of financial resources, and infrastructure. The paper concludes with an emphasis on the need for a strategic plan for m-learning in higher education in Sudan.

Keywords: Mobile Learning Policy. Plan Implementation. Educational Institutions. Sudan

1. Introduction

This study examines the state of mobile learning in higher education in Sudan. Qualitative methods have used to collect and analyze the data. Semi-structured interviews have been used to gather the data from educational technology departments in several universities (six peoples has been interviewed), the data has interpreted and main findings has reported. This study could be needed to shed light on the employment of mobile learning in higher education in Sudan as a global orientatation.

The idea of extending the use of mobile devices in educational contexts is flexible and can be applied broadly (Eltayeb & Hegazi, 2014). Therefore, how such technology could be used to improve the education? Mobile learning (m-learning) defined as “learning that can take place in locations” (Savill, Attewell, & Tribal, 2006, p. 11). Furthermore, m-learning is a way of facilitating the process of learning and can be viewed as the process of learning anywhere anytime (Chu & Cai, 2015).

This paper investigates whether Sudan as one of the least developed country with multilingual characteristics and trying to use technology in education to increase the availability of knowledge and digital skills. The country has produced a five-year’s plan (2012-2016) to apply information technology in all educational stages. This plan does not include use of m-learning, as a result, the study investigates whether higher education institutions are planning to use or implement m-learning.

2. Literature Review

The literature of m-learning in higher education addressed several issues as instance policy adoption (Khan et al., 2015), perceptions and challenges of m-learning implementation (Khan et al., 2015), m-learning as tools to support teaching and learning (So, 2016), a task-interaction to support educational decision-making (Fulantelli, Taibi, & Arrigo, 2015), as well as investigations of the student learning mobility and the needs to enable location based learning (David et al., 2017).

In sub-Saharan Africa, still issues of m-learning (Kaliisa & Picard, 2017) is not separate from the context of Sudan, where the lack of literature, however, some studies investigate using mobile phone to develop some skills for kids such as mathematic skills, using Tablet for out of the schoolchildren (Stubbé et al., 2017), and some studies addressed the social part of mobile phone such as uses of m-phones in public places (Khattab & love, 2009).
Nevertheless, using m-learning still in the infancy stages in Sudan in terms of Policies, infrastructure, cultural and role of m-learning, as well as student motivation and skills to use m-learning (Altayb et al., 2014). To conclude that, there is a needs to Pedagogical, critical perspective to implement these technologies and embedded it to the curricula to best serve educational progress.

3. Research Purpose

The purpose of this paper is to find out whether there is planning for m-learning in higher education in Sudan. In the case of positive response, the study would investigate what strategies they are using for the implementation of their plans. In the case of negative response, the study would attempt to find out the factors impeding the employment of m-learning in higher education in Sudan. Based on the purposes of the study, the researcher has proposed these following questions to be answered:

- Is there any institutional strategic plan for mobile learning in higher education in Sudan?
- What are the factors that hinder the use of mobile learning in Sudanese universities?

4. Methodology

The aim of this study is to investigate the status of m-learning in higher education in Sudan and whether educational technology departments have planned or are planning to apply this technology. Qualitative approach has been used to collect the data. The study also has designed and employed semi-structured interviews with several administrators of educational technology departments in different universities to get comprehensive data for the purpose of the study (Bryman, 2008). In other words, the authors conducted an interview with the directors of educational technology departments. The interview has sent to six educational technology departments in six universities, including the Open University of Sudan (OUS), as well as two experts of educational technology.

5. Findings and Discussion

The paper presents the future of m-learning based on the interviewees' prospects, and how it can shape the future of m-learning in Sudan. As mentioned previously, Semi-structured interviews have conducted to collect the data, the interviews were transcribed for subsequent analysis. The study participant’s names have coded (A, B, C), after review of the coded segments led to the emergence of three major themes related to the topic of the study.

5.1 Mobile Policy Adoption

The study investigates the needs to the policy to implement m-learning, the interviewees; however, it is important to planning for mobile technology in Sudanese higher education. In addition to the lack of equipment, there are still other factors impeded the application of m-learning in Sudan. These factors represented in the lack of m-learning policy, financial resources and infrastructure, as well as less considering of the role of m-learning. Notable, there is a policy for ICT which called Five-year’s Plan 2012-2016. Although this policy emphasis on the use of ICT in education, however, it does not include use of m-learning. As a result, the appeal is for formulating policy for m-learning. This step may match with the UNESCO orientation of m-learning “new policy related to mobile learning should be established within existing ICT in education policies which many governments already have in place” (UNESCO, 2013, p. 30).

Addition to the factors above, interviewees A, B, and C believe that “educational technology administrators need to find a way to play effective role in the university planning how to use technology as a whole, and mobile device as part of this technology, the interviewees also emphasized that we need to best utilize of m-learning, raise awareness of educational leaders about the m-learning, in order to promulgate an appropriate policy to support using such kind of technology”. This statement emphasizes
the importance of policy for m-learning because investigations about an institutional plan for using such technology has led to finding out that there is no such plan at all.

5.2 Mobile Usage & Accessibility of Internet

Concerning use and accessibility of internet, mobile devices in particular m-phone is available among teaching staff and students, however, their main use is for social interactions (Facebook, WhatsApp, etc.). This might imply that students may use these social media for learning, therefore, they need to be skilled digitally (how to use m-learning effectively). The interviewees (A, B, and C) were stated the following: It shall need to provide classroom with mobile devices to engage the student in a real learning situation, nonetheless, this step requires the university enrich the student learning and improve teacher’s digital skills. To enable such practice, it is necessary to establish the requirements of m-learning (infrastructure, digital curricula, etc.)

5.3 Future Directions

According to the study participants, higher education in Sudan needs to adopt and implement m-learning requirements from the ground (policy, infrastructure, digital skills, internet availability, and so on) to catchup the regional and global progress in this era, with following the orientation of UNESCO, African development bank, and other good experiences of m-learning.

6. Conclusion

The paper focused on of m-learning adoption in higher education in Sudan in term of strategic plan and policy formulation. Factors affecting m-learning in Sudan has defined and recommendations has been suggested. There is a need to formulate policy, establishing infrastructure, offer technical support, and increase the culture of m-learning to enable this technology in higher education in Sudan.

References

Close Reading of Science Texts with Online Annotations

Guillaume SCHILTZa*, Sarah FRÉDÉRICKXb & Norman SIEROKAa,c

aDepartment of Physics, ETH Zurich, Switzerland
bDepartment of Health Sciences and Technology, ETH Zurich, Switzerland
cDepartment of Humanities, Social and Political Sciences, ETH Zurich, Switzerland

*schiltz@phys.ethz.ch

Abstract: Making annotations is a major technique to investigate the structure and the central ideas of a complex text. In this paper we present an intervention, where undergraduate physics students have been using online annotations to identify and discuss argumentative problems in science texts. Based on the positive results we are planning to apply online annotations to a greater extent.

Keywords: online annotations, close reading, critical thinking, hypothes.is, physics, philosophy

1. Introduction

Online annotations, i.e. annotations directly added to electronic texts, have been pioneered since the early 2000s (Glover Xu and Hardaker, 2007). Those annotations nowadays include highlighting of the original text, glossing of text passages and adding discussion threads directly to the original text. Furthermore, the annotations can be private, public or restricted to a defined group of users. Sharing annotations with peers and teachers is one of the major assets of online annotations (Jones, 2014). Compared to threaded discussion forums, the comments made via online annotations turn out to be more context specific and lead to a more focused discussion (Sun and Gao, 2014). When closely analyzing a text, online annotations, thus, seem to be the preferred method.

A large body of research on learning benefits resulting from online annotations in different educational areas is now available, e.g. Kennedy (2016) for literary studies, Jensen and Scharff (2014) for philosophy, and Tseng, Yeh and Yang (2015) for foreign language acquisition.

With the introduction of online annotations, we wanted to teach and to promote close reading of science texts, especially texts dealing with the philosophy of science. In the past, we have noticed that a great number of our students are lacking techniques to accurately deal with complex argumentative texts. According to Brummett (2010) close reading involves “the mindful, disciplined reading of an object (i.e. text) with the view to a deeper understanding of its meaning”. The main approach of close reading consists in determining which argumentative claims are the most important and how they fit together to support the author’s main ideas. Close text reading, especially focusing on non-narrative complex science texts, supports students to engage more easily in critical thinking and problem solving while developing their communication and collaboration skills (Cummins, 2013, Lapp et al., 2013).

Johnson, Archibald and Tenenbaum (2010) already reported on the benefits of online annotations to foster general reading comprehension skills. Our aim was to check if these findings are consistent within a close reading setting and if our students are willing to adapt online annotations as part of their learning techniques.

2. Instructional Setting

Within the physics curriculum at ETH Zurich we are offering a first-year elective course “Philosophical Reflections on Physics II” which critically evaluates topics from electrodynamics against a broader historical and philosophical/systematic background. As a major learning objective of this course,
students should be able to critically evaluate different topics and approaches in physics. They should also be enabled to communicate their insights to peers from other disciplines and fields. Part of the weekly assignment consists in reading and discussing online an original science text related to the topic that is to be covered in-class. In the past, we have noticed that students experience difficulties in accurately reading these texts. For this reason, in 2017, we have introduced an extra two-week module focusing on close reading. Part of this module is a web annotation tool added to an instructional unit on close reading. Two philosophical science texts were made available in the annotation tool for training purposes. During the two-week run of the module, students were invited to use the annotation tool and to submit at least one online annotation or a reply to an existing annotation.

We have chosen hypothes.is (https://hypothes.is) as annotation tool (Figure 1). Hypothes.is is free of charge and offers an easy to use interface with flexible options for private, public and group annotations. The annotation system can easily be embedded in a WordPress (https://wordpress.org) or Moodle (https://moodle.org) environment. It is developed and maintained by a non-profit organization and has a clear commitment to educational purposes.

The course included 6 plenary in-class sessions as well as 6 monitored face-to-face discussion meetings in fixed groups (5-11 participants). In the first week, students got an online introduction to close reading and had to apply close reading to a selected text via hypothes.is. The task was kept rather simple in the way that students should identify problematic argumentative statements and explain their choice. In addition, students were invited to comment on annotations provided by other students. The following discussion meeting focused on the annotations and comments that students had provided online. We repeated this procedure a couple of weeks later with a second close-reading text.

![Figure 1. Screenshot of the annotation tool hypothes.is (right) embedded in a WordPress environment (left).](image)

### 3. Results and Discussion

29 out of 32 students made use of the annotation tool and submitted a total of 68 annotations plus 21 replies to existing annotations. We asked the students if they consider online annotations helpful to understand the texts. 58% answered positively, while 17% did experience only a marginal comprehension gain, the remaining 25% were undecided. Online annotations as a tool to work with texts was highly appreciated (Table 1). All students agreed that online annotations simplify discussions related to texts. With an average count of 58 words, students provided a rather substantial body of annotations and the instructors noticed a considerable increase of the discussion quality (Table 1). To sum up, students have adapted online annotations as a valuable tool for reading and discussing texts, and they have largely met the requirements linked to close reading.

---

**Table 1. Students' feedback on the use of online annotations**

<table>
<thead>
<tr>
<th>Annotation Use</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>29</td>
<td>58%</td>
</tr>
<tr>
<td>Marginal</td>
<td>17</td>
<td>17%</td>
</tr>
<tr>
<td>Undecided</td>
<td>8</td>
<td>25%</td>
</tr>
</tbody>
</table>

**Table 2. Average words per annotation**

<table>
<thead>
<tr>
<th>Average Words</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>58</td>
<td></td>
</tr>
</tbody>
</table>
We decided to adopt this group of the discussion meetings to the close reading module and configured the annotation tool to make annotations visible only at the group level. Public annotations, i.e. annotations visible to the whole class, were available as well. 28% of the annotations were put in the public area, probably by negligence. Group annotations and especially the engendered discussions turned out to be more appropriate for our setting.

Table 1: Selected comments from students and instructors concerning online annotations (translated from German).

| Students: hypothes.is is an awesome tool that facilitates the interaction with text comments substantially. |
| Instructors: Students reported that online annotations are extremely helpful to prepare the discussion meetings and we as instructors were positively surprised by the quality and nature of the annotations. |

According to data privacy regulations, we were not allowed to use the annotations for grading purposes. The third-party organization hypothes.is stores the annotations on external servers, which are out of our control. This violates our legal directions concerning archiving and maintaining students’ performance data. Furthermore, students have to register for hypothes.is by providing a valid email address. With regard to personal data protection, we cannot force students to enter this registration process. Those legal restrictions considerably curtailed our options and we could offer the online annotation functionality only as an optional feature. Students, however, did not feel worried about these concerns. Nevertheless, we have to work out a legal solution if we want to use online annotations at a larger scale. We are presently engaged in negotiations with hypothes.is.

4. Conclusion

The results from our pilot intervention on promoting close reading with an online annotation tool look promising. Students have adopted the tool and were able to meet the instructional goals at a very satisfying level. In the future, we are planning to extend the use of online annotations and to study the learning benefits of online annotations related to close reading in more detail.

References

A Secular Trend Analysis of the Effects of Using ICT in University Education

Yasuhiko TSUJIa, Rieko INABAb, Mieko TAKAHIRAc & Mana TAGUCHId
aFaculty of Liberal Arts, The Open University of Japan, Japan
bFaculty of Liberal Arts, Tsuda University, Japan
cCollege of Humanities, Chubu University, Japan
dCenter for the Promotion of Excellence in Higher Education, Kyoto University, Japan
*tsuji@ouj.ac.jp

Abstract: Promoting education through information and communication technology in universities is considered an important issue. So far, in order to gain an understanding of its actual conditions, institutions such as the Open University of Japan, Kyoto University, and Academic eXchange for Information Environment and Strategy have conducted questionnaire surveys targeting higher education institutions throughout Japan. This paper focuses on the effects of using ICT in education and compares results over time, while also attempting to extract comparable factors from each fiscal year with regard to the details of these effects. We reveal the characteristics of institutions that demonstrated improvement in their educational methods and effects in the 2015 fiscal year survey results.

Keywords: ICT in education, e-learning, higher education, complete enumeration

1. Introduction

Since globalization and open education are increasingly advocated and promoted within the Japanese higher education system, there is a growing demand for a qualitative change in university education to encourage self-directed learning by students, such as active learning. To realize such a transition, promoting education that uses information and communication technology (ICT) in universities is considered a key issue. So far, in order to gain an understanding of its actual conditions, institutions such as the Open University of Japan (2011), Kyoto University (2014), and Academic eXchange for Information Environment and Strategy (AXIES, 2016) have conducted questionnaire surveys targeting higher education institutions throughout Japan. The purpose of these surveys is to comprehensively show the spread of ICT utilization in higher education in Japan and to clarify the characteristics of our country in the world by comparing with similar overseas surveys such as Green (2016). Ultimately, these surveys aim to clarify the factors for promoting ICT utilization education and to propose measures for accelerating the spread of future ICT utilization education at each institution. On the other hand, a secular trend analysis, which involves investigating the differences in the actual conditions during each fiscal year, has not yet been conducted. Furthermore, its causality, that is, determining what factors bring about the effectiveness of ICT use in education, has not been made clear.

Therefore, this paper focuses on the “effects of using ICT” in education and compares their secular trends, while also attempting to extract comparable factors pertaining to the details of such effects by fiscal year (FY). Additionally, it reports on our investigation concerning the characteristics of institutions that demonstrated “improvement of educational methods” and “educational effects” in FY 2015 survey results.

2. Comparison of Secular Trends for the Effectiveness of ICT in Education

With regard to the effectiveness or ineffectiveness of introducing ICT in education, Figure 1 indicates the responses gained from the four surveys conducted between FYs 2009 and 2015. The four surveys used the same questions and responses were marked along a four-level scale with regard to its effectiveness/ineffectiveness or as “unsure.” Figure 1 shows that approximately half of the institutions considered ICT to be “effective” in all surveys. On the other hand, institutions that were “unsure” of its
effects increased from 37.8% in 2009 to 43.6% in 2015. This shows that while introducing and using ICT in education has certain positive effects, they have not been sufficiently tested.

![Graph showing effectiveness of ICT education over years](image)

**Figure 1. The Effectiveness of ICT Education.**

3. Extraction of Common Factors based on the Comparison of Secular Trends

With regard to institutions that gave responses other than “ineffective” or “unsure,” we attempted to extract common factors based on the comparison of secular trends with specific data on their effectiveness. Furthermore, with regard to specific survey items, there were 13 items for FYs 2009 and 2010 and 17 items for FYs 2013 and 2015 because four items concerning advanced educational initiatives, such as active learning and problem-based learning (PBL), were added. We then conducted an exploratory factor analysis using maximum likelihood estimation and promax rotation for each response from the surveys of FYs 2013 and 2015 and compared the factors and lower-order items obtained. This resulted in finding the same factor structure in the response data for both FYs, excluding one item. Therefore, we eliminated two items with factor loadings under 0.30 in either of the FYs and one item that showed different factors with high loading that varied from year to year and then calculated the scale score for each factor. Table 1 indicates the common factor items extracted as a result. The factor loadings indicated in the table are based on the result of the exploratory factor analysis of FY 2015 responses. Similar to the study by Tsuji et al. (2016), which also conducted factor analysis of FY 2015 responses, we extracted four factors, including “increasing the university’s brand power,” “improving educational methods,” “educational effects,” and “cost reduction” and confirmed sufficient reliability for each factor using Cronbach’s α coefficient.

### Table 1: Common factors based on the comparison of secular trends.

<table>
<thead>
<tr>
<th>Survey items</th>
<th>α coefficient</th>
<th>Increasing the university’s brand power</th>
<th>Improving educational methods</th>
<th>Educational effects</th>
<th>Cost reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attracted more applicants and international students</td>
<td>.87</td>
<td>.07</td>
<td>-.10</td>
<td>-.07</td>
<td>.00</td>
</tr>
<tr>
<td>Enhanced competitiveness and reputation</td>
<td>.78</td>
<td>.09</td>
<td>.03</td>
<td>-.05</td>
<td>.10</td>
</tr>
<tr>
<td>Expanded range of target students</td>
<td>.74</td>
<td>.05</td>
<td>.05</td>
<td>-.07</td>
<td>.10</td>
</tr>
<tr>
<td>Secured diverse teaching staff</td>
<td>.60</td>
<td>-.02</td>
<td>-.02</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Increased active learning style courses</td>
<td>-.02</td>
<td>1.01</td>
<td>-.04</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td>Increased PBL style courses</td>
<td>.00</td>
<td>.89</td>
<td>-.05</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td>Leveraged useful external teaching materials and contents</td>
<td>.80</td>
<td>.13</td>
<td>.44</td>
<td>.03</td>
<td>.18</td>
</tr>
<tr>
<td>Improved access to learning resources for students outside of the university</td>
<td>-.06</td>
<td>.39</td>
<td>.06</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Enhanced effectiveness of student learning</td>
<td>-.06</td>
<td>-.07</td>
<td>.97</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Enhanced student motivation for learning</td>
<td>-.03</td>
<td>-.01</td>
<td>.93</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td>Improved completion rate of students</td>
<td>.21</td>
<td>.00</td>
<td>.47</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Improved quality of education</td>
<td>.04</td>
<td>.29</td>
<td>.45</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Improved working efficiency of teaching staff</td>
<td>-.07</td>
<td>.01</td>
<td>-.05</td>
<td>.81</td>
<td></td>
</tr>
<tr>
<td>Reduced budget costs</td>
<td>.06</td>
<td>-.02</td>
<td>.01</td>
<td>.69</td>
<td></td>
</tr>
</tbody>
</table>
4. Scale Scores for Factors “Improving Educational Methods” and “Educational Effects”

From Table 1, we considered “improving educational methods” and “educational effects” to be two factors related to the improvement of education and educational quality at universities. Therefore, for FY 2015 survey results (AXIES, 2016), we calculated subscale scores of the two factors by averaging and performed a t-test analysis to examine their correlation to other survey items and score the difference between the choices (yes, no). The results are indicated in Table 2 and Table 3. From these, we observe that universities working on “improving educational methods” tend to mention ICT in education-related initiatives in their action plans, obtain more competitive external funds, and make more progress in introducing and using LMS. Furthermore, they also tend to have university-wide organizations for its promotion as well as organizations that provide technological and educational support.

Table 2: Correlation between the two factors.

<table>
<thead>
<tr>
<th>Survey items</th>
<th>Improving educational methods</th>
<th>Educational effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT is mentioned in the university's vision and action plan</td>
<td>.312**</td>
<td>.138</td>
</tr>
<tr>
<td>Funding for ICT in education</td>
<td>.342**</td>
<td>.355**</td>
</tr>
<tr>
<td>Degree of awareness concerning OER</td>
<td>.400**</td>
<td>.232**</td>
</tr>
<tr>
<td>State of OER provision</td>
<td>.210**</td>
<td>.165**</td>
</tr>
<tr>
<td>State of OER use</td>
<td>.270**</td>
<td>.208*</td>
</tr>
<tr>
<td>State of MOOC provision</td>
<td>.266**</td>
<td>.162</td>
</tr>
</tbody>
</table>

** p<.01, * p<.05

Table 3: Subscale scores of other survey items

<table>
<thead>
<tr>
<th>Survey items</th>
<th>Choice</th>
<th>Improving educational methods</th>
<th>Educational effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of a university-wide promotion organization</td>
<td>Yes</td>
<td>3.08**</td>
<td>3.16*</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2.48**</td>
<td>2.91**</td>
</tr>
<tr>
<td>Obtaining competitive external funds</td>
<td>Yes</td>
<td>2.93**</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2.61**</td>
<td>3.00</td>
</tr>
<tr>
<td>Measuring the effectiveness of ICT in education program</td>
<td>Yes</td>
<td>2.44**</td>
<td>2.74****</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2.84**</td>
<td>3.10***</td>
</tr>
<tr>
<td>Introducing and using a university-wide LMS program</td>
<td>Yes</td>
<td>2.87***</td>
<td>3.07*</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2.31***</td>
<td>2.83*</td>
</tr>
<tr>
<td>Presence of a university-wide technical support organization</td>
<td>Yes</td>
<td>2.89***</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2.33***</td>
<td>2.95</td>
</tr>
<tr>
<td>Presence of a university-wide educational support organization</td>
<td>Yes</td>
<td>2.89***</td>
<td>3.09*</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2.53***</td>
<td>2.91*</td>
</tr>
</tbody>
</table>

*** p<.001, ** p<.01, * p<.05, + p<.1

References


The Role of a Technology and a Classroom Activity for Improving EFL Learners' Oral Performance

Kae NAKAYA
Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Japan
knakaya@mr.hum.titech.ac.jp

Abstract: The final goal of my research is to develop novel methods that support English as a foreign language (EFL) learners and to increase their English-speaking skills. In this paper, I describe the outline of a mobile application for self-study English-speaking, and also show a learning design which combines practice with the mobile application and a classroom pair work. I also discuss the role of both a technology and a classroom activity for improving English as a foreign language (EFL) learners’ oral performance.

Keywords: Computer assisted language learning, oral task, pair work, reflection, oral performance

1. Introduction

The goal of my research is to develop novel methods that support English as a foreign language (EFL) learners and to increase their English-speaking skills. To achieve this goal, I have conducted two researches: (1) I have developed a Mobile Application for the Scaffolded summary speaking Task (MAST) in order to support learners practicing English-speaking by self-study and evaluated the effectiveness of the application on learners’ oral fluency; and (2) I have proposed a learning design which combines self-study using MAST and a reflection activity in pairs in an English class, and have clarified the role of MAST and the pair work for improving learners’ oral performance.

2. A Mobile Application for the Scaffolded Summary Speaking Task (MAST)

Murano (2007) stated the effectiveness of self-study summarizing method for improving oral fluency by using a speech production model (Levelt, 1989), which describes how people process information when speaking. Summarizing tasks offer learners what they would say (preverbal message) and what kind of words they would use (retrieval words from Lexicon), therefore the learners can concentrate on grammatical encoding, which might result in improving fluency. However, the effectiveness has not been investigated by comparing it with the method which does not promote grammatical encoding. In addition, learners used this method only once a month (Muranoi, 2007).

Therefore we developed MAST and evaluated the effectiveness in detail. The learning process using MAST includes the summarizing task and scaffolding practices. A learner reads an English newspaper article that is shown on the screen of MAST and explains its summary. After reading the article, and before summarizing it aloud, the learner conducts the scaffolding practices.

The main scaffolding practice is a short question and answer practice. In the practice, the learner needs to answer questions that are related to main points of the summary. MAST vocalizes the question, the learner answers it, and then MAST relays the sample answer and offers the next question, like a pseudo-interactive conversation with a virtual tutor. By repeating the step, the learner can both clarify the main points of the summary and modify her or his summary, and retain their motivation.

We have conducted an experiment to observe the effectiveness of MAST on learners’ oral fluency by comparing oral fluency scores with the scores of the control group, which was offered only the opportunity of reading aloud tasks that did not focus on grammatical encoding.
The analysis revealed that MAST was effective for Repair fluency that was measured by counting repetitions. The participants using MAST were offered hints for summary points and words so that they were facilitated in grammatical encoding. As a result, they might have been able to process information grammatically faster when speaking, therefore repetitions for their speech became less frequent.

Although we found that MAST could improve learners’ oral fluency in the former research, learners need more help to reflect on their oral output. According to an interview with the participants who used MAST in the experiment, they tended to reflect on their speech in terms of fluency only. Swain (2005) indicated the importance of peers’ feedback to test the listener’s comprehension. Therefore, we need to combine the learning of MAST and activities with peers.

3. A Learning Design Combining Self-study with MAST and a Reflection in Pairs

In my later research, I proposed a learning design that combines self-study with MAST and a reflection activity in pairs in an English class. In this design, learners practice English summary speaking by self-study using MAST, and after that they expound upon its summary to their peers and reflect on the summary in pairs, using a worksheet. The main purpose of my latter research is to clarify the role of technologies and the role of activities in a class in detail.

Using the worksheet of the learning design, a pair of learners conducted work reflection in the following processes. First, a learner offers a summary, which the learner practiced at home using MAST. During the speech, the peer listener draws a picture that shows what the peer could understand. Secondly, the speaker and the listener reflect on the speech. They check what the listener could understand and what she or he could not understand based on the picture. After that they discuss and write down other words, or phrases, that assisted the listener in comprehending more fully.

To observe the effectiveness, I conducted a similar experiment for the duration of four weeks and compared oral performance scores of this learning design with the scores of a preliminary experiment which offered neither scaffolding in MAST nor the paired reflection activity in a class. The analysis results revealed that the learning design was effective for improving learners’ Structural Complexity (non-repeated words per the Analysis of Speech Unit). This result revealed that participants could add more words in one speech unit to express what they wanted to say.

As for the role of reflection activity, learners might gain strategies of how to search and use approximate words, or phrases, when they did not think of appropriate words, or phrases, to express what they wanted to say. In the reflection activity, the learners discussed better expression to the peer, which leads to reflection on concrete experience from a different perspective and construction of more abstract knowledge (Kolb, 1994). Therefore, learners might stretch the retrieval words from Lexicon.

In terms of the role of technologies for self-study, the scaffolding of MAST also has an important role in gaining such a strategy, because the scaffolding might develop readiness by activating knowledge regarding vocabularies that are related to the article. In other words, the scaffolding could lead to greater effectiveness with respect to gaining the strategies for accessing the Lexicon.

In future studies, I have to elaborate the learning design in order to facilitate learners in using the strategies acquired from reflection to the next learning activities. That might lead to more elaborate learning, such as self-regulated learning.

References

Author Index

AGAPITO, Jenilyn 736
ALICBUSAN, Roscoe Nealle 68
ALMEIDA, Vitor 727
ALVES, Marcelo Henrique Monier Junior 662,727
ANDERSON, Tom 914
ANDRES, Juan Miguel L. 62
ANZAI, Yasuhiro 421
AOKI, Terumasa 493
ARAUJO, Thiago 481
ARCUS, Kerri 632
ARUNOPRAYOCH, Nuttaphat 228
ASAI, Yasutaka 688
ASBURY, Elizabeth 632
AWWAL, Nafisa 146
AYEDOUN, Emmanuel 835
AZALEA, Alia 946
AZAMI, Satoko 765
AZCARRAGA, Judith 98
BAGHAEI, Nilufar 397,721,990
BANAWAN, Michelle 162
BARMADA, Bashar 397
BECKING, Dominic 222
BEDE, Felipe 481
BO, Wenjin 996
CAI, Su 558,566
CALEON, Imelda 457
CARE, Esther 146
CARVALHO, Lucila 192
CARVALHO, Zuila 481
CAVALCANTE, Daniel 481
CHAI, Ching Sing 826,1001
CHAI, Wasan Na 80
CHAN, Tak-Wai 246
CHANG, Ben 336
CHANG, Maiga 808
CHE-LEAH, Muttakeen 774
CHEN, Der-Thanq 996
CHEN, Gaowei 539
CHEN, Guang 576,638
CHEN, Gwo-Dong 784
CHEN, Hsiao-Hua 823
CHEN, Ling 427
CHEN, Nian-Shing 240
CHEN, Peng 820
CHEN, Weiqin 373
CHEN, Weni 139
CHEN, Zhi-Hong 246,829
<table>
<thead>
<tr>
<th>Name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAO, Yuan</td>
<td>296</td>
</tr>
<tr>
<td>GATINHO, Fátima</td>
<td>490,662</td>
</tr>
<tr>
<td>GENG, Gretchen</td>
<td>614,620</td>
</tr>
<tr>
<td>GENG, Jie</td>
<td>586</td>
</tr>
<tr>
<td>GEORGE, Jade Li</td>
<td>755</td>
</tr>
<tr>
<td>GODA, Yoshiko</td>
<td>922</td>
</tr>
<tr>
<td>GODHE, Anna-Lena</td>
<td>845</td>
</tr>
<tr>
<td>GOSSELIN, David</td>
<td>210</td>
</tr>
<tr>
<td>GOTODA, Naka</td>
<td>415,451</td>
</tr>
<tr>
<td>GOTTIPATI, Swapna</td>
<td>261,280,391</td>
</tr>
<tr>
<td>GU, Xiaqing</td>
<td>116,403,463,475,502,505</td>
</tr>
<tr>
<td>GUO, Siao-Cing</td>
<td>928</td>
</tr>
<tr>
<td>GWEK, Dina Heng Li</td>
<td>280</td>
</tr>
<tr>
<td>HÄKKINEN, Päivi</td>
<td>146</td>
</tr>
<tr>
<td>HAMMEDI, Sanda</td>
<td>808</td>
</tr>
<tr>
<td>HANNA, Junko</td>
<td>922</td>
</tr>
<tr>
<td>HASEGAWA, Shinobu</td>
<td>252</td>
</tr>
<tr>
<td>HATONO, Isuo</td>
<td>549</td>
</tr>
<tr>
<td>HAYASHI, Yuki</td>
<td>21,74,107,316,835</td>
</tr>
<tr>
<td>HAYASHI, Yusuke</td>
<td>11</td>
</tr>
<tr>
<td>HAYASHIDA, Yukuo</td>
<td>712</td>
</tr>
<tr>
<td>HECKING, Tobias</td>
<td>326</td>
</tr>
<tr>
<td>HERRADURA, Tita</td>
<td>113</td>
</tr>
<tr>
<td>HEW, Khe Foon</td>
<td>539</td>
</tr>
<tr>
<td>HIRASHIMA, Tsukasa</td>
<td>11,445</td>
</tr>
<tr>
<td>HOEL, Tore</td>
<td>373</td>
</tr>
<tr>
<td>HONG, Hun-Gi</td>
<td>496</td>
</tr>
<tr>
<td>HOPPE, H. Ulrich</td>
<td>326,1007</td>
</tr>
<tr>
<td>HORI, Masahiro</td>
<td>955</td>
</tr>
<tr>
<td>HSIAO, I-Han</td>
<td>427</td>
</tr>
<tr>
<td>HSU, Chung-Yuan</td>
<td>826</td>
</tr>
<tr>
<td>HSU, Ting-Chia</td>
<td>928</td>
</tr>
<tr>
<td>HU, Cheng-Hao</td>
<td>925</td>
</tr>
<tr>
<td>HU, Menghua</td>
<td>403,730</td>
</tr>
<tr>
<td>HU, Yiling</td>
<td>463,730</td>
</tr>
<tr>
<td>HU, Yi-Ling</td>
<td>237</td>
</tr>
<tr>
<td>HUANG, Nen-Fu</td>
<td>469</td>
</tr>
<tr>
<td>HUANG, Ronghuai</td>
<td>638,820</td>
</tr>
<tr>
<td>HUANG, Shihua</td>
<td>367</td>
</tr>
<tr>
<td>HWANG, Gwo-Jen</td>
<td>549</td>
</tr>
<tr>
<td>HWANG, JauJian</td>
<td>549</td>
</tr>
<tr>
<td>IBARRIENTOS, Chester Paul</td>
<td>95</td>
</tr>
<tr>
<td>IHARA, Daiki</td>
<td>385</td>
</tr>
<tr>
<td>IKEDA, Mitsuru</td>
<td>21,204,955</td>
</tr>
<tr>
<td>INABA, Rieko</td>
<td>1039</td>
</tr>
<tr>
<td>INAGAKI, Shigenori</td>
<td>715,790,796</td>
</tr>
<tr>
<td>INAZAWA, Keisuke</td>
<td>345</td>
</tr>
</tbody>
</table>
ISHIGE, Yumi 922
ISHIMURA, Tsukasa 511
ISHIZUE, Ryosuke 40
ITOH, Yukihiro 104,270,288,385,421
IWANE, Noriyuki 487
JAN, Mingfong 832
JIA, Shan 367,558
JOHNSON, Andrew 896
JULIA, Chen 884
JONATHAN, Christin 457
JONG, Siu Yung Morris 586
KALAW, Kristine 98
KANEHIRA, Ren 1030
KANEKO, Keiichi 733
KAO, Gloria Yi-Ming 249
KARUNGARU, Stephen 415,451
KASHIHARA, Akihiro 345
KAWAGUCHI, Shuya 796
KAWANO, Hirokazu 243
KHAMBARI, Mas Nida 594
Md. KIM, Jaewook 694
KINSUK 240
KISHI, Yasuhito 694
KNIGHT, Simon 210,306
KOGURE, Satoru 104,288,385,421
KOH, Elizabeth 457,517
KOH, Joyce 1001
KOJIRI, Tomoko 50
KONDO, Makoto 104
KONG, Siu-Cheung 984
KONISHI, Tatsuhiro 104,270,288,385,421
KONNO, Satoru 765
KONO, Takanori 499
KONOMI, Shinichi 433,439,514,682,908
KOTANI, Katsunori 934
KOTHIYAL, Aditi 166
KOZAKI, Shun 712
KUMAMOTO, Etsuko 549
KUO, Meng-Lun 469
KUO, Tonny Meng-Lun 216
KURAYAMA, Megumi 110
KUSUNOKI, Fusako 715,790,796
LAI, ChihHung 228
LAKSHMI T G 974
LAM, Vincent 884
LAO, ChanChio 984
LAPINID, Minie Rose 68
LEE, Alwyn Vwen Yen 119
LEE, Chang Youn 496
LEE, Chia-An 469
<table>
<thead>
<tr>
<th>Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>MITSUISHI, Takashi</td>
<td>493</td>
</tr>
<tr>
<td>MIWA, Honomi</td>
<td>955</td>
</tr>
<tr>
<td>MIWA, Kazuhisa</td>
<td>92</td>
</tr>
<tr>
<td>MIZOGUCHI, Hiroshi</td>
<td>790,796</td>
</tr>
<tr>
<td>MIZUTANI, Kozo</td>
<td>520</td>
</tr>
<tr>
<td>MONIER, Elza</td>
<td>490,662</td>
</tr>
<tr>
<td>MONIER, Marcelo</td>
<td>676</td>
</tr>
<tr>
<td>MORI, Natsumi</td>
<td>74</td>
</tr>
<tr>
<td>MORITA, Naoya</td>
<td>508</td>
</tr>
<tr>
<td>MORITA, Yusuke</td>
<td>931</td>
</tr>
<tr>
<td>MOSES, Priscilla</td>
<td>946</td>
</tr>
<tr>
<td>MOURI, Kousuke</td>
<td>529,733</td>
</tr>
<tr>
<td>MUANGON, Weenawadee</td>
<td>890</td>
</tr>
<tr>
<td>MUROTA, Masao</td>
<td>855</td>
</tr>
<tr>
<td>MUROYA, Daiki</td>
<td>107</td>
</tr>
<tr>
<td>MURTHY, Sahana</td>
<td>166,974</td>
</tr>
<tr>
<td>MUSARRAT, Raghib</td>
<td>802</td>
</tr>
<tr>
<td>NAGAHAMA, Toru</td>
<td>931</td>
</tr>
<tr>
<td>NAGAO, Katashi</td>
<td>508</td>
</tr>
<tr>
<td>NAKAGAWA, Seiichi</td>
<td>902</td>
</tr>
<tr>
<td>NAKAJIMA, Ai</td>
<td>865</td>
</tr>
<tr>
<td>NAKANO, Midori</td>
<td>487</td>
</tr>
<tr>
<td>NAKAYA, Kae</td>
<td>855,1042</td>
</tr>
<tr>
<td>NARAYANA, Soumya</td>
<td>974</td>
</tr>
<tr>
<td>NARITA, Yosuke</td>
<td>31</td>
</tr>
<tr>
<td>NATARAJAN, Uma</td>
<td>1001</td>
</tr>
<tr>
<td>NATE, Fumito</td>
<td>50</td>
</tr>
<tr>
<td>NEHRING, Natalia</td>
<td>721</td>
</tr>
<tr>
<td>NESHATIAN, Kourosh</td>
<td>604</td>
</tr>
<tr>
<td>NING, Lu</td>
<td>280</td>
</tr>
<tr>
<td>NISHIZAKI, Hiromitsu</td>
<td>31</td>
</tr>
<tr>
<td>NOGAMI, Tomoyuki</td>
<td>715</td>
</tr>
<tr>
<td>NOGUCHI, Yasuhiro</td>
<td>104,270,288,385,421</td>
</tr>
<tr>
<td>NORIKO, Uosaki</td>
<td>733</td>
</tr>
<tr>
<td>OCHARO, Harriet</td>
<td>252</td>
</tr>
<tr>
<td>OGAN, Amy</td>
<td>1</td>
</tr>
<tr>
<td>OGATA, Hiroaki</td>
<td>529,908</td>
</tr>
<tr>
<td>OHASHI, Hiroshi</td>
<td>908</td>
</tr>
<tr>
<td>OHIRA, Shigeki</td>
<td>508</td>
</tr>
<tr>
<td>OI, Misato</td>
<td>682</td>
</tr>
<tr>
<td>OKAMOTO, Masaru</td>
<td>487,499,511</td>
</tr>
<tr>
<td>OKAZAKI, Keita</td>
<td>56</td>
</tr>
<tr>
<td>OKAZAKI, Yasuhisa</td>
<td>355,712</td>
</tr>
<tr>
<td>OKUBO, Fumiya</td>
<td>439,514,682</td>
</tr>
<tr>
<td>OKUGAWA, Kaito</td>
<td>104</td>
</tr>
<tr>
<td>OLIVEIRA, Ana Emilia</td>
<td>409,490,662</td>
</tr>
<tr>
<td>OLIVEIRA, Francisco</td>
<td>481</td>
</tr>
<tr>
<td>ONG, Ethel</td>
<td>68,98,95</td>
</tr>
</tbody>
</table>
ONO, Yuichi 865
ORSBORN, Georgina 632
OSTROWSKI, P. Christopher 129
OUYANG, Shuxian 576
OWEN, Helen 755
PAGUILIGAN, Moira Denise 95
PALADINES, José 101
PARNELL, Roderic 210
PENNINGTON, Deana 210
PENUGONDA, Harshavardhan 974
PERRIS, Kirk 1033
PHATTALUNG, Ratthakarn Na 774
PONGPINIGPINYO, Sunee 890
POON, Chung Keung 86
POONIA, Vishvendra 974
POWERS, David 914
PÖYSÄ-TARHONEN, Johanna 146
Qi, Cong 176
RABELO, Aldrea 409,490,662,676
RABELO, Dilson 409,662
RAMÍREZ, Jaime 101
RAMLI, Murni 1001
REDDY, Lou 990
REDDY, Y Raghu 802
REYNOLDS, Barry 914
ROBSON, Daphne 626
RODRIGO, Ma. Mercedes 1,62,186,198,234,718,736
ROFIE, Fadzil Saleh Mohamad 594
ROMERO, Margarida 817
RONGHUAI, Huang 1033
RONGSHENG, Jeff Lin 261
RUANGRAJITPAKORN, Taneth 80
RUI, Song 280
RYAN, Philippa 306
SAKAMOTO, Kazunori 40
SAMPAIO, Higo 676,727
SANCHEZ, Byron 493
SCHILTZ, Guillaume 1036
SETA, Kazuhsia 21,74,107,316,835
SHAN, Jun-Hao 1018
SHANKARARAMAN, Venky 261,391
SHARMA, Vaishali 802
SHI, Rui 475,502
SHIBANI, Antonette 306
SHIH, Yen-An 336
SHIMADA, Atsushi 433,439,514,733
SHISHIBORI, Masami 746
SHUM, Simon Buckingham 156,306
SIROKA, Norman 1036
SILVA, Francisco 727
SILVA, Mauro 676
SMITH, Adam 896
SOARES, Maikon 481
SONG, Yanjie 671
SOONKLANG, Tasanawan 890
SPINILLO, Carla 409
SRIJUNTONGSIRI, Gun 204
STEINMEIER, Christine 222
SUAREZ, Merlin Teodosia 113
SUBRAMANIAN, Shankar 724
SUGAWARA, Kohta 415,451
SUGIMOTO, Aoi 316
SUGIMOTO, Masanori 790,796
SUGIMOTO, Sirak Kaewjamnong 890
SUGIYAMA, Takumi 288
SUN, Daner 656
SUN, Jian 566
SUN, Yuting 638
SUPNITHI, Thepchai 80,204
SUSELO, Thomas 361
SUZUKI, Yubun 908
SVOBODA-GOUVEA, Julia 210
SWANSI, Vanlalhruaii 113
TABATA, Yoshiyuki 549
TACHINO, Takashi 694
TAGUCHI, Mana 1039
TAHA, Mohannad 1033
TAIRAB, Adam 1033
TAKAHIRA, Mieko 1039
TAKEDA, Yoshiaki 796
TAN, Aik Ling 937
TAN, Jennifer PeiLing 457,517
TAN, Seng Chee 119,937
TANAKA, Koji 955
TANAKA, Yui 715
TANG, Chung Man 86
TANG, Ying 539
TANIGUCHI, Yuta 514,682
TAWATSUJI, Yoshimasa 56
TAY, Aaron 280
TAY, Siu Hua 457
TENGCHAISRI, Vachirawit 733
THO, Pham-Duc 228
THOMAS, Michael 832
THOMPSON, Kate 210
TOKUOKA, Mikihiro 790
TOKUTAKE, Keitaro 50
YANG, Kai-Hsiang 814,823
YANG, Lin-Wan 1013
YANG, Wan-Lin 832
YARZEBINSKI, Evelyn 1
YEOMAN, Pippa 192
YIM, Joanne 946
YIN, Baoyuan 367
YIN, Chengjiu 549,733
YOKOYAMA, Yuuki 665
YOSHIKAWA, Atsushi 355
YOSHIKAWA, Takehiko 415,451
YOSHIMI, Takehiko 934
YU, Fu-Yun 1024,1027
YU, Liang-Chih 919
YU, Yuen Tak 86
ZHANG, Lishan 379,427
ZHANG, Qiujie 427
ZHANG, Xiumin 493
ZHENG, Longwei 475,502,505
ZHOU, Peng 650
ZHUANG, Zijun 379
ORGANISED BY APSCE
Asia-Pacific Society for Computers in Education

SPONSORS

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

UC ENGINEERING

Unitec Institute of Technology
TE WHARE WĀNANGA O MAIRAEA

Allied Telesis