Rethinking Learning in the Digital Age. Making the Learning Sciences Count
Volume 1

Judy Kay and Rosemary Luckin
Rethinking Learning in the Digital Age: Making the Learning Sciences Count
13th International Conference of the Learning Sciences (ICLS) 2018

Volume 1

13th International Conference of the Learning Sciences

June 23-27, 2018, London, United Kingdom
University College London (UCL)

Editors:
Judy Kay and Rosemary Luckin
Conference Organizers and Committees

Conference Co-Chairs
Kaska Porayska-Pomsta, UCL Knowledge Lab, UK
Mavrikis, Manolis, UCL Knowledge Lab, UK
(with Kristen Weartherby, UCL Knowledge Lab, UK and Suklaa Ltd)

Program Co-Chairs
Judy Kay, University of Sydney, Australia
Rose Luckin, UCL Knowledge Lab, UK

Chair of the Conference Committee for ISLS
Oskar Lindwall, University of Gothenburg, Sweden

Advisory Committee
Oskar Lindwall, University of Gothenburg, Sweden
Brian Smith, Drexel University, USA
Carolyn P. Rose, Carnegie Mellon University, USA
Chee-Kit Looi, Nanyang Technological University, Singapore
Cindy Hmelo-Silver, Indiana University, USA
Frank Fischer, University of Munich, Germany
Kris Lund, University of Lyon, France
Nikol Rummel, Ruhr-University Bochum, Germany
Peter Reimann, University of Sydney, Australia
Susan Yoon, Penn Graduate School of Education, USA

Workshop Co-Chairs
Amalia Cristina Urdiales Garcia, University of Malaga, Spain
Iris Tabak, Ben Gurion University, Israel

Doctoral Consortium Co-Chairs
Benedict du Boulay, University of Sussex, UK
Carolyn P. Rose, Carnegie Mellon University, USA

Early/Mid Career Workshop Co-Chairs
Eduardo Guzmán De los Riscos, University of Malaga, Spain
Erica Halverson, University of Wisconsin, USA

Industrial, Commercial Co-Chairs
Vania Dimitrova, University of Leeds, UK
Wayne Holmes, Open University, Milton Keynes, UK

Practitioner Co-Chairs
Mutlu Cukorova, UCL Knowledge Lab, UK
Jane Hunter, Technical University Sydney, Australia

Communications Chair
Manolis Mavrikis, UCL Knowledge Lab, UK
(with Anissa Moeni and Veronica Cucuiat, Suklaa Ltd, Education Media Centre, London)
Senior Reviewers

Shaaron Ainsworth
Christa Asterhan
Michael Baker
Evrin Baran
Brian Belland
Yifat Ben-David Kolikant
Dani Ben-Zvi
Orit Ben-Zvi Assaraf
Matthew Berland
Gautam Biswas
Daniel Bodemer
Marcela Borge
Susan Bridges
Paul Bma
Angela Calabrese Barton
Carol Chan
Tak-Wai Chan
Elizabeth Charles
Wenli Chen
Ching-Huei Chen
Clark Chinn
Ming Chiu
Doug Clark
Alison Clark-Wilson
Tamara Clegg
Ulrike Cress
Crina Damşa
Joshua Danish
Ton de Jong
Carrie Demmans Epp
Benjamin DeVane
Yannis Dimitriadis
Vania Dimitrova
Betsy DiSalvo
Benedict du Boulay
Ravit Golan Duncan
Julia Eberle
Deborah Fields
Frank Fischer
Barry Fishman
Nobuko Fujita
Dragan Gasevic
Aristotelis Gkiolmas
Soon Joo Gog
Susan Goldman
Art Graesser
Monique Grandbastien
Kai Hakkarainen
Erica Halverson
Jiangang Hao
Katie Headrick Taylor
Dan Hickey
Cindy Hmelo Silver
Yotam Hod
Mike Horn
Michael Jacobson
Isa Jahnke
Heisawn Jeong
Yasmin Kafai
Yael Kali
Judy Kay
Yoon Jeon "YJ" Kim
Paul Kirschner dr.h.c.
Simon Knight
Ingo Kollar
Janet Kolodner
Janet Kolodner
Eleni Kyza
Nancy Law
Heather Leary
Victor Lee
Sharon Levy
Lijia Lin
Oskar Lindwall
Lucia Linn
Chee-Kit Looi
Kristine Lund
Johan Lundin
Lina Markauskaite
Steven McGee
Eva Millán Valldoperas
Tanja Mitrovic
Gaelle Molinari
Yishay Mor
Johannes Moskaliuk
Mitchell Nathan
Omid Noroozi
Amy Ogan
Chandra Orrill
Jun Oshima
Annemarie Palincsar
Zachary Pardos

Teresa Pargman
Frederick Peck
Bill Penuel
Joseph Polman
Michael Prilla
Mimi Recker
Peter Reimann
Ido Roll
Carolyn Rosé
Nikol Rummel
William A. Sandovall
Eileen Scanlon
Tammy Schellens
Baruch Schwarz
Mamta Shah
Kshitij Sharma
Jarek Sierschynski
Sergey Sosnovsky
Florence Sullivan
Kate Thompson
Mike Tissenbaum
Dimitra Tsouvaltzi
Philip Vahey
Jan van Aalst
Wouter Van Joolingen
Keisha Varma
Freydis Vogel
Hanna Vuojarvi
Erin Walker
Susan Walker
Barbara Wasson
Joanna Weidler-Lewis
Armin Weinberger
David Weintrop
Michelle Wilkerson
Philip Winne
NiAll Winters
Alyssa Wise
Lung Hsiang Wong
Ruth Wylie
Kalina Yacef
Susan Yoon
Diego Zapata-Rivera
Gustavo Zurita Alarcón
**Reviewers**

<table>
<thead>
<tr>
<th>Anthony Aakre</th>
<th>Sylvie Barma</th>
<th>Murat Cakir</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotem Abdu</td>
<td>Jackie Barnes</td>
<td>Angela Calabrese Barton</td>
</tr>
<tr>
<td>Dor Abrahamson</td>
<td>Lauren Barth-Cohen</td>
<td>Regina Calloway</td>
</tr>
<tr>
<td>Robin Adams</td>
<td>Lauren Barth-Cohen</td>
<td>Ingrid Cao</td>
</tr>
<tr>
<td>Jennifer Adams</td>
<td>Sarit Barzilai</td>
<td>Mia Carapina</td>
</tr>
<tr>
<td>Lalitha Agnihotri</td>
<td>Satabdi Basu</td>
<td>Spencer Carlson</td>
</tr>
<tr>
<td>Shirley Agostinho</td>
<td>Kathryn Bateman</td>
<td>Cynthia Carter Ching</td>
</tr>
<tr>
<td>June Ahn</td>
<td>Daniel Belenky</td>
<td>Paulo Carvalho</td>
</tr>
<tr>
<td>Shaaron Ainsworth</td>
<td>Adam Bell</td>
<td>Moraima Castro-Faix</td>
</tr>
<tr>
<td>Anna Åkerfeldt</td>
<td>Brian Bellland</td>
<td>Veronica Caver</td>
</tr>
<tr>
<td>Fabio Akhras</td>
<td>Anat Ben Shabat</td>
<td>Dionne Champion</td>
</tr>
<tr>
<td>Samir AKNINE</td>
<td>Mordechai Ben-Ari</td>
<td>Margaret Cava</td>
</tr>
<tr>
<td>Carlos Alario-Hoyos</td>
<td>Hava Ben-Horin</td>
<td>Kennedy Chan</td>
</tr>
<tr>
<td>Hamed Alavi</td>
<td>Nina Bergdahl</td>
<td>Lap Ki Chan</td>
</tr>
<tr>
<td>Tugce Aldamir</td>
<td>Peter Bergstrom</td>
<td>Margaret Chan</td>
</tr>
<tr>
<td>Vincent Aleven</td>
<td>Jonte Bernhard</td>
<td>Peichin Chang</td>
</tr>
<tr>
<td>Sakina Alhadad</td>
<td>Marina Bers</td>
<td>Ching-Yu Chang</td>
</tr>
<tr>
<td>Laura Allen</td>
<td>Dario Bertero</td>
<td>Ben Chang</td>
</tr>
<tr>
<td>Carrie Allen</td>
<td>Bronwyn Bevan</td>
<td>Yu-Hui Chang</td>
</tr>
<tr>
<td>Danielle Allessio</td>
<td>Ruchi Bhanot</td>
<td>Selem Charfi</td>
</tr>
<tr>
<td>Alicia Alonzo</td>
<td>Xiangfang Bi</td>
<td>Koula Charitonos</td>
</tr>
<tr>
<td>Sarah ALSalaijman</td>
<td>Katerine Bielaczy</td>
<td>Elizabeth Charles</td>
</tr>
<tr>
<td>Tugba Altan</td>
<td>Harm Biemans</td>
<td>Bernadette Charlier</td>
</tr>
<tr>
<td>Isabel Alvarez</td>
<td>John Biggam</td>
<td>Clement Chau</td>
</tr>
<tr>
<td>Sagaya Amalathas</td>
<td>Jose Blackorby</td>
<td>Muhammad Ali Chaudhry</td>
</tr>
<tr>
<td>Elise Ameloot</td>
<td>Paulo Blikstein</td>
<td>Ching-Huei Chen</td>
</tr>
<tr>
<td>Ricardo Amorim</td>
<td>Daniel Bodemer</td>
<td>Grace Chen</td>
</tr>
<tr>
<td>Stamatina Anastopoulou</td>
<td>Rebecca Boncoddio</td>
<td>Qing Chen</td>
</tr>
<tr>
<td>Mona Anchan</td>
<td>elizabeth bonsignore</td>
<td>Bodong Chen</td>
</tr>
<tr>
<td>Emma Anderson</td>
<td>Marcela Borge</td>
<td>Zan Chen</td>
</tr>
<tr>
<td>Christopher Andrews</td>
<td>Nigel Bosch</td>
<td>Guanhua Chen</td>
</tr>
<tr>
<td>Jerry Andriessen</td>
<td>Ivica Botički</td>
<td>Ying-Chih Chen</td>
</tr>
<tr>
<td>Gabriella Anton</td>
<td>François Bouchet</td>
<td>Gaowei Chen</td>
</tr>
<tr>
<td>Lauren Applebaum</td>
<td>Helen Bound</td>
<td>Li Cheng</td>
</tr>
<tr>
<td>Béatrice Arend</td>
<td>Anne Boyer</td>
<td>San Chye Cheng</td>
</tr>
<tr>
<td>Nancy Ares</td>
<td>Corey Brady</td>
<td>John Cherniavsky</td>
</tr>
<tr>
<td>Maria Aristeidou</td>
<td>Tharrenos Bratstis</td>
<td>Ellina Chernobilsky</td>
</tr>
<tr>
<td>Ivon Arroyo</td>
<td>Bert Bredeweg</td>
<td>Cuye Cheshire</td>
</tr>
<tr>
<td>Tultaleni I. Asino</td>
<td>Karen Brennan</td>
<td>Joshua Childs</td>
</tr>
<tr>
<td>Christa Asterhan</td>
<td>Susan Bridges</td>
<td>Doris Chin</td>
</tr>
<tr>
<td>N'a'sma Av-Shalom</td>
<td>Torsten Brinda</td>
<td>Tan Ying Chin</td>
</tr>
<tr>
<td>Guat Poh Aw</td>
<td>Liat Brix-Etgar</td>
<td>Clark Chinn</td>
</tr>
<tr>
<td>Roger Azevedo</td>
<td>Christopher Brooks</td>
<td>Jennifer Chiu</td>
</tr>
<tr>
<td>Afaa Baabdullah</td>
<td>Rachael Brown</td>
<td>Thomas K.F. Chiu</td>
</tr>
<tr>
<td>Ari Bader-Natal</td>
<td>orit broza</td>
<td>Jennifer Chiu</td>
</tr>
<tr>
<td>Haesol Bae</td>
<td>Emma Brunskill</td>
<td>Kyungjin Cho</td>
</tr>
<tr>
<td>Connor Bain</td>
<td>Sarah Buchanan</td>
<td>Gi Woong Choi</td>
</tr>
<tr>
<td>Michael Baker</td>
<td>Juergen Buder</td>
<td>Irene-Angelica Chounta</td>
</tr>
<tr>
<td>Ryan Baker</td>
<td>James Burke</td>
<td>Doris Choy</td>
</tr>
<tr>
<td>Arthur Bakker PhD</td>
<td>Winslow Burleson</td>
<td>Köppe Christian</td>
</tr>
<tr>
<td>Megan Bang</td>
<td>Jim Bywater</td>
<td>Man-Wai Chu</td>
</tr>
<tr>
<td>Amanda Barany</td>
<td>Donna Caccamise</td>
<td>Stefanie Chye</td>
</tr>
<tr>
<td>Wolmet Barendregt</td>
<td>Huining Cai</td>
<td>Matthieu Cisel</td>
</tr>
</tbody>
</table>
Preface

The International Conference of the Learning Sciences (ICLS) is a major international event, organized biennially by the International Society of the Learning Sciences (ISLS): a professional society dedicated to the interdisciplinary empirical investigation of learning as it exists in real-world settings and to how learning may be facilitated both with and without technology. The international and interdisciplinary field of the Learning Sciences brings together researchers from the fields of cognitive science, educational research, psychology, computer science, artificial intelligence, information sciences, anthropology, sociology, neurosciences, and other fields to study learning in a wide variety of formal and informal contexts (see www.isls.org). The field emerged in the late 1980s and early 1990s, with the first ICLS held in 1991 at Northwestern University in Evanston, Illinois, USA. Subsequent meetings of ICLS were held again in Evanston, USA (1996), Atlanta, GA, USA (1998), Ann Arbor, MI, USA (2000), Seattle, WA, USA (2002), Santa Monica, CA, USA (2004), Bloomington, IN, USA (2006), Utrecht, the Netherlands (2008), Chicago, IL, USA (2010), Sydney, NSW, Australia (2012), and Boulder, CO, USA (2014) and National Institute of Education, Nanyang Technological University of Singapore (2016).

A bumper crop of 965 submissions for ICLS 2018 were received in December 2017 (571 were received at the last ICLS Conference in 2016). All submissions went through a process of rigorous peer review. The acceptance rate was 32% for full papers, 27% for short papers and 33% for posters. 110 of the full papers and 61 of the short papers were accepted as posters. For symposia, the acceptance rate was also 32%. At least three reviewers were recruited for each paper and a member of the Programme Committee provided a metareview. The Programme Chairs made decisions, based on reading every review, metareview and the online discussion.

This meant that over 3000 reviews were completed by the ICLS community – many, many thanks. We are particularly indebted to the committee members and reviewers who responded to the call for volunteers to manage extra papers. We especially note those who managed at least 9 papers and those reviewers who responded so generously to calls for additional reviews: we could not have completed the selection process without you.

The UCL Institute of Education, London, is hosting the 13th International Conference of the Learning Sciences (ICLS), 2018, from June 23rd to 27th, 2018. It is a right and fitting venue to host the 13th International Conference of the Learning Sciences (ICLS). UCL Institute of Education (IOE) was founded in 1902, and is a world-leading centre for research and teaching in social science and education. For three successive years (2014, 2015, 2016 and 2017), the Institute has been ranked first for education worldwide in the QS World University Rankings, and was shortlisted for the ‘University of the Year’ title in the 2014 Times Higher Education (THE) awards. In January 2014, Ofsted judged it to be ‘outstanding’ in every category for initial teacher training across primary, secondary and further education programmes. In the most recent Research Excellence Framework, 94% of the IOE’s research was judged to be world class. In 2016, it was awarded the Queen’s Anniversary Prize for Higher and Further Education, for its innovative social research and contribution to policy and practice in education. The IOE currently has more than 8,000 students, 800 staff, and attracts students from over a hundred countries around the world. Since December 2014, it has been a school of University College London, formally called the UCL Institute of Education. University College London (UCL) was founded in 1826, and was the first English university established after Oxford and Cambridge, the first to admit students regardless of race, class, religion or gender, and the first to provide systematic teaching of law, architecture and medicine. We are among the world's top universities, as reflected by performance in a range of international rankings and tables. UCL currently has over 35,000 students from 150 countries and over 11,000 employees.

The theme for this year’s conference is Rethinking learning in the digital age: Making the Learning Sciences Count. This reflects the fact that now, more than ever, the learning sciences have a key role to play in unpacking the complexity of the teaching and learning process. AI and Automation in the workplace, including within education, will alter what we need to learn and how we need to teach it. Therefore, as scientists and educators we need to explore learning in real-world settings in an interdisciplinary manner in order to understand how learning may be facilitated both with and without technology.

For the first time this year we also welcomed Crossover paper submissions to reflect the fact that ICLS 2018 is co-located with L@S and AIED, as part of the London Festival of Learning. Ten papers were accepted (acceptance rate 30%). These papers appeal to a broad audience of researchers from across the communities of the three conferences. The London Festival of Learning is a unique opportunity to bring together world experts in artificial intelligence in education, the learning sciences and technical innovations in education. There has never been a more important time for these three disciplines to meet and overlap, uniting academics to share their research and learn from each other, as well as engaging with a wider audience of educators, businesses and
learners. The Festival of Learning also offers an opportunity to showcase the important work being done by UCL IOE's EDUCATE programme in promoting the best in EdTech development currently taking place in the UK.

We hope you will enjoy what promises to be a week of fascinating debate, discussion and international networking.

Judy Kay, Human Centred Technology Research, University of Sydney, Australia
Rosemary Luckin, UCL Knowledge Lab, UCL Institute of Education, UK
Table of Contents

Volume 1

Keynotes

Digital Footprints and Shoes that Don’t Fit................................................................. Iris Tabak
Devising Ethnographic Eyes: Tracing Learning Processes and Practices Across Multiple Levels of Times,
Space, and Sociocultural Configurations................................................................. Judith Green
Transitioning Education’s Knowledge Infrastructure: Shaping Design or Shouting From the Touchline? Simon Buckingham Shum

Full Papers

Planning to Iterate: Supporting Iterative Practices for Real-world Ill-structured Problem-solving........... Daniel Rees Lewis, Jamie Gorson, Leesha V. Maliakal, Spencer Evan Carlson, Elizabeth Gerber,
Christopher Riesbeck, Matthew Easterday
Negotiating Epistemic Agency and Target Learning Goals: Supporting Coherence from the Students’
Perspective.................................................................................................................. Aliza Zivic, John F. Smith, Brian Reiser, Kelsey Edwards, Michael Novak, Tara McGill
Exploring the Unknown: Supporting Students’ Navigation of Scientific Uncertainty
With Coupled Methodologies....................................................................................... Julia Svoboda Gouvea, Aditi Wagh
Science Engagement and Identities in Everyday Family Life ......................................................... Dana Vedder-Weiss
A Focus on Contribution Towards Product and Performance in Collaborative Design.................. Colin Hennessy Elliott, Sarah C. Radke, Jasmine Y. Ma
Exploring Multimodal Scaffolds Supporting Middle School Students’ Construction of
Causal-Mechanistic Scientific Explanations........................................................................ Nitasha Mathayas, David Brown
Scaffolding Authentic Scientific Inquiry Experiences for Early Elementary Learners Using Wearable
Technology.................................................................................................................. Virginia L. Byrne, Seokbin Kang, Leyla Norooz, Rafael Velez, Monica Katzen, Afe Addeh, Jon Froehlich,
Tamara Clegg
How Teachers Use Instructional Improvisation to Organize Science Discourse and Learning
in a Mixed Reality Environment.................................................................................. Maggie Dahn, Noel Enyedy, Joshua Danish
Designing for Rightful Presence in 6th Grade Science: Community Ethnography as Pedagogy............ Angela Calabrese Barton, Edna Tan
Designing Technology as a Cultural Broker for Young Children: Challenges and Opportunities........ Yanghee Kim, Sherry Marx, Hung Pham Ngoc, Tung Nguyen
Determinants of School Level Success in Design-Based Innovation Networks............................... Nancy Law, Yeung Lee, Constance Wong-Loke
Informing the Design of Teacher Awareness Tools through Causal Alignment Analysis ................................. 104
Kenneth Holstein, Bruce M. McLaren, Vincent Aleven

The Effect of the Prior Collaborative Experience on the Effectiveness and Efficiency of Collaborative Learning ....................................................................................................................................... 112
Jimmy Zambrano R., Femke Kirschner, Paul Kirschner

Encountering and Becoming Role Models: Combating Underrepresentation in STEM .................................... 120
Leighanna Hinojosa

The Influence of Students’ Cognitive and Motivational Characteristics on Differences in Use and Learning Gain in an E-Learning Environment ..................................................................................... 128
Charlotte Larmuseau, Jan Elen, Piet Desmet, Fien Depaepe

Authentic Problem-Based Learning with Augmented Reality............................................................................ 136
Todd Ogle, David Hicks, Aaron Johnson, Thomas Tucker

Toward a Taxonomy of Team Performance Visualization Tools ....................................................................... 144
Zachari Swiecki, David Williamson Shaffer

Is Scrolling Disrupting While Reading? ............................................................................................................. 152
Katherine Brady, Sun Joo Cho, Gayathri Narasimham, Douglas Fisher, Amanda Goodwin

Exploring Predictors of Secondary School Teachers’ Use of Technology to Support Student-Centered Teaching ................................................................................................................................ 160
Yaoran Li, Vitaliy Popov, Veronica Garza, Anne Keicher

STEAM Learning in an In-school Makerspace: The Role of Distributed Spatial Sensemaking ......................... 168
Kay E. Ramey, Reed Stevens, David Uttal

Secondary Students’ Evaluation of Inappropriate Strategies of Reasoning About Evidence Under a Scientific Explanation ........................................................................................................... 176
Guanzhong Ma, Jan van Adst, Carol Chan, Jing Wang

Flow in Computer-Supported Collaborative Problem-Solving........................................................................... 184
Gaelle Molinari, Sunny Avry

Epistemic Agency as a Members’ Experience.................................................................................................... 192
Danielle Keifert, Christina (Stina) Krist, Deana A. Scipio, Anna McLean Phillips

Learning Nanoscience Concepts Through a Nanoscale Experience ................................................................ 200
Polly K. Lai, Michael Jacobson, Micah Goldwater

Chinese Character Composition Game for Collaborative Language Learning .................................................. 208
Yun Wen, Wenli Chen

Learning From Errors – The Effect of Comparison Prompts in Instruction After Problem Solving Settings ........................................................... 216
Katharina Loibl, Timo Leuders

Identifying Reflective and Non-Reflective Group Consensus Strategies for Evidence-Based Scientific Argumentation ........................................................................................................... 224
Susan Yoon, Miyoung Park, Emma Anderson

Redistributing Epistemic Agency: How Teachers Open Up Space for Meaningful Participation in Science ........................ 232
Mon-Lin Ko, Christina (Stina) Krist

Examining the Role of Explicit Epistemic Reflection in Promoting Students’ Learning from Digital Text ..................................................................................................................................... 240
Feng Lin, Dana Gnesidilow, Nicole D. Martin, Catherine Louise Dornfeld Tissenbaum, Sadhana Puntambekar

Bringing Practices of Co-Design and Making to Basic Education ........................................................................ 248
Sini Riikonen, Pirite Seitamaa-Hakkarainen, Kai Hakkarainen
Trade-offs in Using Mobile Tools to Promote Action with Socioscientific Issues ........................................ 256
Susan Yoon, Jooeun Shim, Noora F. Noushad

Martina A. Rau, Miranda Renee Zahn

A Collaboration Script for Nonverbal Communication Enhances Perceptual Fluency with Visual Representations...................................................................................................................... 272
Martina A. Rau, Purav Patel

Teachers’ Mediation of Students’ Interactions with Physical and Virtual Scientific Models in Biology........ 280
Nicole D. Martin, Dana Gnesdilow, Sadhana Puntambekar

Designing for and Analyzing Productive Uncertainty in Science Investigations ........................................ 288
Eve Manz

Public Peer Review Motivates Higher Quality Feedback .................................................................................. 296
Xu Wang, Yali Chen, Amanda Godley, Carolyn Rosé

The Design and Evaluation of Optimal Computerized Guidance for Invention Activities: The Invention Coach ............................................................................................................................... 304
Catherine Chase, Helena Connolly, Marianna Lamnina, Vincent Aleven

Matching Data-Driven Models of Group Interactions to Video Analysis of Collaborative Problem Solving on Tablet Computers .................................................................................................................. 312
Luc Paquette, Nigel Bosch, Emma Mercier, Jiyoung Jung, Saadeddine Shehab, Yurui Tong

Conjecture Mapping the Library: Iterative Refinements Toward Supporting Maker Learning Activities in Small Community Spaces ........................................................................................................................................ 320
Victor R. Lee, Mimi Recker, Abigail Leigh Phillips

Blocks or Text? How Programming Language Modality Makes a Difference in Assessing Underrepresented Populations .......................................................................................................................... 328
David Weintrop, Heather Killen, Baker E. Franke

Towards a Cognitive Ecological Framework in CSCL .......................................................................................... 336
Marcela Borge, Emma Mercier

Youth, Learning and Social Media in K-12 Education: The State of the Field .................................................... 344
Emilia Askari, Diana Brandon, Sarah Galvin, Christine Greenhow

Defining and Assessing Risk Analysis: The Key to Strategic Iteration in Real-World ........................................ 352
Spencer Evan Carlson, Leesha V Maliakal, Daniel Rees Lewis, Jamie Gorson, Elizabeth Gerber, Matthew Easterday

A Study of the Design and Enactment of Scientific Modeling Tasks to Support Fourth-Grade Students’ Sense-Making ..................................................................................................................................... 360
Meredith Baker Marcum

Examining Pre-service Teacher Knowledge Trajectories of Computational Thinking Through a Redesigned Educational Technology Course ........................................................................................................ 368
Hui Yang, Chrystalla Mouza, Yi-Cheng Pan

Interconnecting Knowledge, Experience, and Self in Humanistic Knowledge Building Communities ............. 376
Yotam Hod, Dani Ben-Zvi, Michal Dvir

A Principled Approach to Designing Assessments That Integrate Science and Computational Thinking ......... 384
Satabdi Basu, Kevin W McElhaney, Shuchi Grover, Christopher J. Harris, Gautam Biswas

Investigating the Coupling of Narrative and Locality in Augmented Reality Educational Activities: Effects on Students’ Immersion and Learning Gains ........................................................................................................ 392
Collaborative Gesture as a Case of Distributed Mathematical Cognition .......................................................... 552  
Candace A. Walkington, Geoffrey Chelule, Dawn Woods, Mitchell Nathan

Professional Development for Secondary Science Teachers: A Faded Scaffolding Approach to Preparing Teachers to Integrate Computing........................................................................................................................ 560  
Colby Tofel-Grehl, Kristin A. Searle, David F. Feldon

Measuring the Scale Outcomes of Curriculum Materials ................................................................................... 568  
Natalie Pareja Roblin, Chris Schunn, Susan McKenney

Developing Assessments that Measure Core Ideas and Scientific Practices: Challenges and Insights .............. 576  
Veronica L. Cavera, Ravit Golan Duncan, Clark Chinn, Moraima Castro-Faix

In the Hive: Designing for Emergence When Teaching Complex Systems In Early Childhood.......................... 584  
Kylie A. Peppler, Naomi Thompson, Joshua Danish, Armin Moczek

Becoming Facilitators of Creative Computing in Out-of-School Settings ......................................................... 592  
Ricarose Roque, Rupal Jain

A Climate of Support: A Process-Oriented Analysis of the Impact of Rapport on Peer Tutoring ...................... 600  
Michael Madaio, Kan Peng, Amy Ogan, Justine Cassell

Augmenting Qualitative Analyses of Collaborative Learning Groups Through Multi-Modal Sensing ............... 608  
Bin Xie, Joseph M. Reilly, Yong Li Dich, Bertrand Schneider

Multimodal Texts and Tasks in Elementary Project-based Science ................................................................... 616  
Miranda S. Fitzgerald, Annemarie Palincsar

“Ohhh, Now I Can Do It!”: School-age Children’s Spontaneous Mathematical Sensemaking in Construction Play ........................................................................................................................................... 624  
Lara Jasien, Ilana Seidel Horn

Teaching with a Fully Digital, Year-long Math Program: Learning Sciences Futures on the Front Line ............ 632  
Jeremy Roschelle, Phillip Herman, Karen Bumgardner, Nicole Shechtman

Interlacing Gaze and Actions to Explain the Debugging Process ....................................................................... 640  
Kshitij Sharma, Katerina Mangaroska, Michail Giannakos, Pierre Dillenbourg

Teacher Practice Spaces: Examples and Design Considerations ........................................................................ 648  
Justin Reich, Yoon Jeon "J" Kim, Kevin Robinson, Dan Roy, Meredith Thompson

Mapping Networks to Help Education Leaders Gain Insights into Complex Educational Systems .................. 656  
Robbin Riedy, Katie Van Horne, Phillip Bell, Bill Pennell, Tiffany Neill, Sam Shaw

Teachers’ Values in Co-Design of an Art-Science-Computation Unit ............................................................... 663  
Lila Finch, R. Benjamin Shapiro, Franziska Carstens

Volume 2

Full Papers (continued)

A New Approach to Lesson Study Practice in Japan from the DBIR Perspective ............................................. 673  
Jun Oshima, Ritsuko Oshima, Yuji Chiyonishio, Ayano Ohsaki

Innovative Collaborative Learning Practice: An ‘Innovative’ Delphi Approach ............................................... 681  
Rose Luckin, Mutlu Cukurova

Combining Gaze, Dialogue, and Action from a Collaborative Intelligent Tutoring System to Inform Student Learning Processes .................................................................................................................... 689  
Jennifer Olsen, Kshitij Sharma, Vincent Aleven, Nikol Rummel
Cultural Repertoires: Indigenous Youth Creating With Place and Story
Kristin A. Searle, Teresa Casort, Breeanne K. Litts, Bryan McKinley Jones Brayboy, Sequoia Lynn Dance, Yasmin Kafai

Exploring Practices on the Move: Facilitating Learning Across a Neighborhood
Lautaro Cabrera, June Ahn, Jason Yip, Tamara Clegg, Kenna Hernly, Elizabeth Bonsignore, Caroline Pitt, Daniel Paauw

From Theory to Practice: How Pre-service Science Teachers Learn to Become Social Justice Educators
Jarod Kawasaki, Kathryn Anderson-Levitt, Linsdey Nenadal, Nadine Tanio, Josephine Pham, Claudia Diera, Annamarie Francois

Design Considerations for Capturing Computational Thinking Practices in High School Students’ Electronic Textile Portfolios
Debora Lui, Gayithri Jayathirtha, Deborah Fields, Mia Shaw, Yasmin Kafai

Revising Biology Misconceptions Using Retrieval Practice and Explanation Prompts
Merrin Oliver, Maggie Renken, Joseph Jay Williams

Using Iterative Design to Create Efficacy-Building Social Experiences with a Teachable Robot
Nichola Lubold, Erin Walker, Heather Pon-Barry, Yuliana Flores, Amy Ogan

Rethinking Loafers: Understanding the Productive Functions of Off-Task Talk During Collaborative Mathematics Problem-Solving
Jennifer Langer-Osuna, Emma Gargroetzi, Rosa Chavez, Jen Munson

Reassembling Home-work: Mixing “Newer” and “Older” Technologies in Home Learning Environments
Deborah Silvis, Katie Headrick Taylor

Patterns of Classroom Talk Through Participation in Discourse-Focused Teacher Professional Development
William A. Sandoval, Alexander J. Kwako, Anahid S. Modrek, Jarod Kawasaki

Changes in Students’ Use of Epistemic Criteria in Model Evaluation
Na’ama Y. Av-Shalom, Hebbah El-Moslimany, Ravit Golan Duncan, Clark Chinn

Enhancing Reflective Learning Experiences in Museums Through Interactive Installations
Hai Huang, Wei Hong Lo, Kher Hui Ng, Timothy Brailsford, Claire O’Malley

Co-Framing Shared Epistemic Objects of Inquiry to Support Knowledge Building Over a Whole School Year
Dan Tao, Jianwei Zhang, Dandan Gao

Recognizing Competencies vs. Completion vs. Participation: Ideal Roles for Web-Enabled Digital Credentials
Daniel T. Hickey, Grant Chartrand

Solder and Wire or Needle and Thread: Examining the Effects of Electronic Textile Construction Kits on Girls’ Attitudes Towards Computing and Arts
Richard Lee Davis, Chris Proctor, Michelle Friend, Paulo Blikstein

Do Alternative Instructional Approaches Result in Different Learning Progressions?
Moraima Castro-Faix, Amber Todd, William Romine, Ravit Golan Duncan

Encouraging Revision of Scientific Ideas with Critique in an Online Genetics Unit
Emily Jean Harrison, Libby Gerard, Marcia Linn

Designing Dialogues on Writing and College Readiness Across Educational Institutions
Alecia Marie Magnifico, Christina Ortmeier-Hooper

Power in the Digital Age: A Critical Revision to Productive Disciplinary Engagement (PDE)
Priyanka Agarwal, Tesha Sengupta-Irving
Design Matters: The Impact of Technology Design on Students’ Inquiry Behaviors ................................................. 839
Engin Bumbacher, Zahid Hossain, Ingmar Riedel-Kruse, Paulo Blikstein

Engineering Discourse Development in an Informal Youth-Driven Maker Club ..................................................... 847
Sagit Betser, Lee M. Martin

The Influence of Students’ Transformative and non-Transformative Contributions on Their Problem Solving in Collaborative Inquiry Learning ......................................................................................................... 855
Vitaliy Popov, Wendi Xing, Gaoxia Zhu, Paul Horwitz, Cynthia McIntyre

Rethinking the Teaching and Learning of Area Measurement .................................................................................. 863
Nicole Panorkou

Characterizing Computational Thinking in High School Science .................................................................................. 871
Hillary Swanson, Golnaz Arastoopour Irgens, Connor Bain, Kevin Hall, Philip Woods, Carson Rogge, Mike Horn, Uri Wilensky

Developing Productive Discourse among Low Achievers in a Knowledge Building Environment ........................ 879
Yuyao Tong, Carol Chan, Jan van Aalst

“We Were on the Same Level”: Young Engineering Researchers Taking Up Agentive Positions in a Diverse Learning Community .................................................................................................................. 887
Wendy Wakefield, Michelle E. Jordan, Mia DeLaRosa

Knowledge Building Inquiry and Reflection in Developing Children's Epistemology of Science ............................ 895
Carol Chan, Cindy Xu, Feng Lin

Short Papers
Building a Team Leadership Index Through Computational Methods ................................................................. 905
Kui Xie, Gennaro Di Tosto, Lin Lu

Equity-oriented STEM-rich Making Among Youth From Historically Marginalized Communities .................... 909
Angela Calabrese Barton, Edna Tan

"Your Turn!": Playing Cooperative Modern Board Games to Promote Perspective Taking and Cooperative Attitudes .................................................................................................................. 913
Yu-Chi Wang, Jennifer Doll, Keisha Varma

Playing and Designing Games for Systems Thinking: A Design Based Research Project .................................... 917
Amanda M. Bell

Forms of Emergent Collaboration in Maker-Based Learning ....................................................................................... 921
Erica Halverson, Breanne K. Litts, Brian Gravel

Combining Non-Programming Activities with Programming for Introducing Foundational Computing Concepts .................................................................................................................. 925
Shuchi Grover, Nicholas Jackiw, Patrik Lundh, Satabdi Basu

Preschool-Age Children Practicing Science: The Intersection of Explanations, Modeling, and Gesture Use ....... 929
Julia Plummer, Amy Ricketts

Dynamic Exploration on Self-explanation Prompts in Complex Tasks ................................................................. 933
Hyun Joo, Jinju Lee, Dongsik Kim

Drawing for Learning from Dynamic Visualizations in Science .............................................................................. 937
Mike Stieff, Katharina Scheiter, Shaaron Ainsworth, Claudia Bohrmann-Linde, Max Schall

Making Learning Journeys Visible: Towards Supporting Collective Reflection on Graduate Attributes ............ 941
Roberto Martinez-Maldonado, Theresa Dirndorfer Anderson, Iván Silva Feraud, Simon Buckingham-Shum

Researcher or Fellow Citizen? Looking for a Role Model in the Humanities .............................................................. 945
Acceptance and Refusal: Examining Conflicting Goals Within Co-Design ................................................................. 1025
Kristina Stamatis

The Quality of Open Online Learning and Education: Towards a Quality Reference Framework for MOOCs ..................................................................................................................................................... 1029
Christian M. Stracke, Esther Tan

Taking on the Challenges of Learning in the Digital Age: Grade 5 Students’ Mindsets and Strategies in Knowledge Building Communities .................................................................................................................................................. 1033
Yotam Hod, Jianwei Zhang, Guangji Yuan, He Zhou

Evidence-based Reasoning of Pre-Service Teachers: A Script Perspective .................................................................. 1037
Katharina Kiemer, Ingo Kollar

Social Positioning Newcomer Roles on a High School Robotics Team: A Chronotopic Micro-Analysis .............. 1041
Colin Hennessy Elliott

Developing a Library of Typical Problems During Collaborative Learning in Online Courses .................................................................................................................................................................................. 1045
Sebastian Strauß, Nikol Rummel, Filipa Stoyanova, Nicole Krämer

Critique Processes in Digital Journalism ........................................................................................................................ 1049
Jessie Nixon

Words Mean Things: How Museum Workers’ Discursive Practices Position the Diverse Communities They Seek to Engage ........................................................................................................................................ 1053
Krystal Villanosa, Mike Horn

Agent-Based Models to Support Bioscience Learning in Nursing Education .......................................................... 1057
Ilana Dubovi, Victor R. Lee

Authentic Learning and Teaching in an Out-of-School Lab - First Steps towards Empirical Investigation of a Theoretical Model ........................................................................................................................................ 1061
Valentina Nachtigall, Angelina Keuschnig, Lena Behrendt, Laura Brune

Beyond Analogy: Qualitative Dimensions of Comparing in Math Class ....................................................................... 1065
Sarah L White

Social Network Analysis for Signaling Pedagogical Shifts in Challenge-Based and Traditional Online STEM Courses ........................................................................................................................................ 1069
Catherine Louise Dornfeld Tissenbaum, Kemi Jona

Sequencing Arithmetic, Area, and Algebraic Instruction for Teaching the Distributive Principle ................................. 1073
Soo-hyun Im, Sashank Varma

Indoor Positioning Technology & Enhanced Engagement in Early Elementary Systems Thinking and Science Learning .................................................................................................................................................. 1077
Kylie A. Pepler, Naomi Thompson, Joshua Danish, Armin Moczek, Shenshen Han

Qualitative Measures of Equity in Small Groups ........................................................................................................... 1081
Benjamin Archibeque, Mary Bridget Kustusch, Florian Genz, Scott Franklin, Eleanor C. Sayre

Bringing Static Code to Life: The Instructional Work of Animating Computer Programs With the Body .......... 1085
Virginia J. Flood, David Deliema, Dor Abrahamson

Third Graders’ Use of Digital Tools Designed for Multimodal Communication in Project-based Science ...... 1089
Miranda S. Fitzgerald, Gabriel P. DellaVecchia, Annemarie Palincsar, Elliot Soloway

Guiding Intent Participation at an Art Crating Company ................................................................................................. 1093
Sarah C. Radke, Jasmine Y. Ma

Using Phenomenography in Educational Technology Research From 2003 to 2017: A Systematic Review and Content Analysis .................................................................................................................................................. 1097
Sally Wai Yan Wan, Sancia Wai-San Wan
Teachers’ Views on Supporting Self-Regulated Learning in Early Childhood Science Education 1101
Audrey Kittredge, Natalie Day, Lenka Janík Blaskova, Heyi Zhang, Sara Baker

Exploring the Margins of the Field: Rethinking STEM in Education 1105
Christopher Peter Ostrowski, Sandra Becker, Zulay Diaz Caceres, Marilu Lam-Herrera, Stefan Rothschuh

Multimodal Learning Analytics for the Qualitative Researcher 1109
Marcelo Worsley

Exploring Teacher Learning through STEM Teachers’ Exploration of Data Using a Domain Specific Coding Language 1113
Detra Price-Dennis, Charles Lang

Conceptual Patterns of Changes in University Students’ Explanations During DC-Circuit Tasks 1117
Terhi Mäntylä, Tommi Kokkonen

“It Didn’t Really Go Very Well”: Epistemological Framing and the Complexity of Interdisciplinary Computing Activities 1121
Serena Thoma, Elise Deitrick, Michelle Wilkerson

Exploring Relevant Problem-Solving Processes in Learning from Productive Failure 1125
Charleen Brand, Christian Hartmann, Nikol Rummel

Reflecting on Epistemic Ideals and Processes: Designing Opportunities for Teachers’ Epistemic Growth 1129
Shiri Mor Hagani, Sarit Barzilai

Dealing with Changes and Challenges: Grade 5 Students’ Experience with Knowledge Building Pedagogy in a Yearlong Science Inquiry 1133
He Zhou, Jianwei Zhang

How Does Expansive Curricular Framing Support Productive Epistemological Framing of Computational Modeling Activities? 1137
Aditi Wagh, Julia Svoboda Gouvea

Missing the Brilliance of Scholars of Color: Mathematics Teacher Educator Discourse in a White Zone of Proximal Development 1141
Ashley D. Scroggins, Victoria Hand, Tonya Bartell, Sunghwan Byun, Beth Herbel-Eisenmann, Courtney Koestler

Co-Navigating Mobilized Student Inquiry Across Multiple Contexts 1145
Ryan Rish, Aijuan Cun

Assessing Equity in Collaborative Learning Situations: A Comparison of Methods 1149
Paul Hutchison, Michael Slattery

Open Discussions of Historical Activity Inquiry Questions 1153
Thérèse Laferrière, Alain Breuleux, Sylvie Barma, Gyeong Mi Heo, Stephanie Beck, Nilou Baradaran, Alice Vanlint

Teachers Collaboratively Creating Micro-Credentials for Professional Development 1157
Sandy Powell, Heather Leary, Lisa McLachlan, Karen Brock

Infrastructuring for Participatory Design of School Technology Practices: How Students Refined Design Practices 1161
Ung-Sang Lee

Design Math: Middle-School Youth Making Math by Building Yurts 1165
Kylie A. Peppler, Mishael Sedas, Tarrence Banks, John Searcy, Scott R. Wallace

Cultivating Formative Intervention Research Partnerships in Mathematics 1169
Charles Munter, Cara Haines, Rebecca Bruton
Developing Theory-Practice Understanding Through Online Discourse Among Pre-Service Teachers .......... 1173
Carol Chan, Kennedy Chan, Ka Lok Cheng

Examining Primary Teacher Expertise and Agency in the Collaborative Design of Project-Based Learning Innovations ........................................................................................................................................ 1177
Samuel Severance, Joseph Krajcik

Towards a Radical Healing Praxis for Black Girls: Imagining Learning Environments That Foster the Sociopolitical Learning of Adolescent Black Girls ........................................................................................................ 1181
Ginnie I. Logan

Symposia
Designing for Axiological Innovation Within Family-Centered Learning Environments .......................... 1187
Enrique Suárez, Carrie Tzou, Megan Bang, Meixi *, Ricarose Roque, Nichole Pinkard, Raymond McDermott, Brigid Barron, Shelley Goldman, Megan Luce, Tanner Vea, Luke Conlin, Philip Bell, Caitlin Kennedy Martin

Video Data and the Learning Event: Four Case Studies ........................................................................... 1195
Ricardo Nemirovsky, Elizabeth de Freitas, Kate O’Brien, Molly L. Kelton, Jasmine Y. Ma, Francesca Ferrara, Giulia Ferrari, Rogers Hall, Lauren Vogelstein, Nathalie Sinclair

Exploring the Adoption, Spread, and Sustainability of an Informal STEAM Learning Innovation In Schools ......................................................................................................................................................... 1203
Reed Stevens, Kay E. Ramey, Peter Meyerhoff, Jaakko Hilppö, Kristiina Kumpulainen, Anu Kajamaa, Antti Rajala, Richard Halverson

The Challenge of Assessing “Knowledge in Use”: Examples from Three-Dimensional Science Learning and Instruction .................................................................................................................................................. 1211
James W. Pellegrino, Brian Douglas Gane, Sania Zahra Zaidi, Christopher J. Harris, Kevin W. McElhaney, Nonye Alozie, Phyllis Haugabook Pennock, Samuel Severance, Knut Neumann, David Fortus, Joe Krajcik, Jeffrey Nordine, Erin Marie Furtak, Derek Briggs, Rajendra Chattergoon, Bill Penuel, Kerri Wingert, Katie Van Horne

Knowledge Analysis Outside the STEM Classroom ...................................................................................... 1219
Eleanor Anderson, Ayush Gupta, Thomas M. Philip, Lina Markauskaite, Yael Kali, Peter Goodyear, Arthur Hjorth, Olivia Levrini

Orchestration Tools for Teachers in the Context of Individual and Collaborative Learning: What Information Do Teachers Need and What Do They Do With It? ................................................................................. 1227
Anouschka van Leeuwen, Nikol Rummel, Kenneth Holstein, Bruce M. McLaren, Vincent Alevin, Inge Molenaar, Carolien Knoop-van Campen, Baruch Schwarz, Naomi Prusk, Osama Swidan, Avi Segal, Kobi Gal

Networked by Design: Interventions for Teachers to Develop Social Capital ................................................. 1235

Moving Forward: In Search of Synergy Across Diverse Views on the Role of Physical Movement in Design for STEM Education ............................................................................................................ 1243
Dor Abrahamson, Alejandro Andrade, Oskar Lindwall, Arthur Bakker, Mitchell Nathan, Candace A. Walkington, Robb Lindgren, David Brown, Asnat R. Zohar, Sharona T. Levy, Joshua Danish, Adam Maltese, Noel Envedy, Megan Humbug, Asmalina Saleh, Maggie Dahn, Christine Lee, Xintian Tu, Bria Davis, Chris Georgen

Unpacking Dimensions of Evidentiary Knowledge and Reasoning in the Teaching and Learning of Science ......................................................................................................................................................... 1251
Posters

Equitable Science Outcomes and School Organizational Conditions
John Settlage

Visual Feedback for Asynchronous Online Interaction: Exploratory Study on the Pattern of Other-regulation
Ji Young Lim, Ji Hyun Lee, Kyu Yon Lim

Dropping in to Game Design: Iterations of a Skatepark Physics Game for a Children’s Museum Exhibit
Benjamin DeVane, Jeremy Dietmeier, Ben J. Miller, Kristen Missall, Salloni Nanda

The Role of Social-academic Goals in Chinese Students’ Self-regulated Learning
Jing Wang, Kun Liu, Guanzhong Ma

Design of a Virtual Internship to Develop Technological Pedagogical Content Knowledge
Diler Oner

Making Energy Easy: Interacting with the Forces Underlying Chemical Bonding Using ELI-Chem Simulation
Asnat R. Zohar, Sharona T. Levy

Elementary Science Preservice Teachers’ Use of Evidence From Rehearsals When Reflecting on Revisions of Their Practice
Anna Maria Arias, Sarah J. Fick, Amanda Benedict-Chambers

Using Machine Learning Techniques to Capture Engineering Design Behaviors
Jim P. Bywater, Mark Floryan, Jennifer Chiu, Jie Chao, Corey Schimpf, Charles Xie, Camilo Vieira, Alejandra Magana, Chandan Dasgupta

Mapping Research and Writing Mentorship Assemblages in a Mixed Cohort Course-based Research Experience
Adam Papendieck, Yin Hong Cheah, Chad Eliason, Julia Clarke

Cueing Gestures in a Seasons Simulation: Outcomes of an Embodied Learning Approach to Supporting Explanations
Robert C. Wallon, Robb Lindgren

The Effects of Inquiry-Based Learning in Higher Education Statistics Tutorials on Students’ Self-Efficacy, Attitudes, and Achievement Emotions
Petra Bod, Daniel Sommerhoff

Competing Epistemologies in the Construction of Popular Science
Pryce Davis

Measuring Integrated Knowledge – A Network Analytical Approach
Marcus Kubisch, Jeffrey Nordine, Knut Neumann, David Fortus, Joe Krajcik

Comparison of 3D Display Technologies for Embodied Interaction in Virtual Hands-On Experiential Learning
Yulong Bian, Qiuchen Wang, Chao Zhou, Guowen Qi, Chenglei Yang, Xiangyu Meng, Chia Shen

Contextual Dimensions of an Ambient Intelligent Classroom
Matthew Montebello

Uncovering the Rich Club Phenomenon in an Online Class
Tianhui Huang, Bodong Chen

Peer Tutor Matching for Introductory Programming: Data-Driven Methods to Enable
New Opportunities for Help ................................................................. 1377
Nicholas Diana, Michael Eagle, John Stamper, Shuchi Grover, Marie Bienkowski, Satabdi Basu

Sketching and Gesturing for New Ideas in Collaborative Design ................................................................. 1379
Tellervo Härkki, Pirta Seittaan-Hakkarainen, Kai Hakkarainen

Developing Assessment Tasks to Promote Student Sensemaking of Phenomena and Flexible Thinking ....... 1381
Emily Miller, Susan Codere, Joe Krajcik

Data Moves: Restructuring Data for Inquiry in a Simulation and Data Analysis Environment ......................... 1383
Michelle Wilkerson, Kathryn Lanouette, Rebecca Shareff, Natalya St. Clair, Nicole Bulalacao, Tim Erickson, Joan I. Heller, William Finzer, Frieda Reichsman

Is Student Frustration in Learning Games More Associated with Game Mechanics or Conceptual Understanding? ......................................................................................................................... 1385
Shamya Karumbaiah, Seyedahmad Rahimi, Ryan Baker, Valerie Shute, Sidney DMello

Transitioning to an Integrated Science Teaching Model: Easier Said than Done .............................................. 1387
Ashley Iveland, Elizabeth B. Dyer, Edward Britton, Burr Tyler, Joshua Valcarcel

The Interaction of the Need for Cognitive Closure With Implicit and Explicit Guidance in Wiki-Based Learning ................................................................................................................................. 1389
Sven Heimbuch, Daniel Bodemer

Research Questions to Support Conversational Learning in the Era of Ubiquitous, Mobile Agents ............ 1391
Bob Schloss, Maria D. Chang, Aditya Vempaty, Arup Acharya, Ravindranath Kokku, Lorin Wilde, Nirmal Mukhi

Middle School Student Ideas on the Relative Affordances of Physical and Virtual Models ......................... 1393
Elizabeth McBride, Jonathan Vitale, Marcia Linn

Detecting Patterns of Dynamic Teacher-Learner Interactions in Online Adult Learning Through a Dynamic Systems Approach ........................................................................................................... 1395
Yohei Kato, Michael Tscholl, Saskia Kunnen

Secondary Students’ Model-Based Reasoning about Earth Systems: Practice, Epistemology, and Conceptual Understanding .................................................................................................................. 1397
Cory Forbes, Mark A. Chandler, Devarati Bhattacharya, Tina Vo, Jane Griffin

Exploring Novice Approach to Conceptual Design of Software .................................................................... 1399
T. G. Lakshmi, Sridhar Iyer

What’s the Difference? A Closer Look at Idea-Centric Analysis of Online Discourse in K12 and Higher Education Settings .................................................................................................................... 1401
Ablyn Vwen Yen Lee, Seng Chee Tan

Enhancing Online Structured Dialogue During Teaching Internships Through Digital Storytelling to Promote Professional Socialization ...................................................................................................... 1403
Toshio Mochizuki, Takeshi Kitazawa, Jun Oshima, Hideyuki Suzuki, Hideo Funaoi

Enskilment in the Digital Age: The Interactional Work of Learning to Debug ............................................. 1405
Virginia J. Flood, David DeLiema, Benedikt Walter Harrer, Dor Abrahamson

Robots That Help: Moving Toward Human-Centered Designs ........................................................................ 1407
Erin Tolar, Andrea Sarah Gomoll, Pei-Jung Li, Benjamin Oistad, Selma Sabanovic, Cindy Hmelo Silver

Child-Coach-Parent Network for Early Literacy Learning ........................................................................... 1409
Juliana Nazare, Anneli Hershman, Ivan Sysoev, Lauren Fratamico, Juanita Butrago, Mina Soltangheis, Sneha Priscilla Makini, Eric Chu, Deb Roy

Learner Expertise and Emotional Design in Multimedia Learning ................................................................ 1411
Thomas K.F. Chiu
Problem Scoping in Designing Biomimetic Robots ................................................................. 1413
Fayette Shaw, Kristen Wendell, Gillian Puttick, Debra Bernstein, Ethan Danahy

Mentor Academy: Engaging Global Learners in the Creation of Data Science Problems for MOOCs .......... 1415
Rebecca M. Quintana, Christopher Brooks, Cinzia Villanucci Smothers, Yuanru Tan, Zheng Yao, Chinmay Kulkarni

The Student-Centric Electronic Portfolio in Practice ................................................................. 1417
Auli Saarinen, Pirita Setiamaa-Hakkarainen, Kai Hakkarainen

Designing Learning Environments to Facilitate Creativity ....................................................... 1419
Jonan Phillip Donaldson

Social Comparison Theory as Applied to MOOC Student Writing: Constructs for Opinion and Ability ...... 1421
Heeryung Choi, Nia Dowell, Christopher Brooks

Can Speaking Make Learning Easier? Verbal Rehearsal Effects on Cognitive Load, Learning Efficacy
and Performance ............................................................................................................................. 1423
Toni M. Hatten-Roberts, Jason M. Lodge

Visualizations of Community Knowledge for Supporting Middle School Students to Model Phenomena in
Scientific Inquiry ............................................................................................................................... 1425
Michelle Lui, Tom Moher, Brenda Lopez Silva

A New Facet: Building Multifaceted Engineering Identity ......................................................... 1427
Jordan O. James, Vanessa Svihta, Chen Qiu, Abhaya K. Datye

Pre-service Teachers’ Perspectives on Computer Science Education Within an Equity-Oriented Teacher
Education Program ......................................................................................................................... 1429
Kelsey Tayne, R. Benjamin Shapiro, A. Susan Jurow, Max Hollingsworth

Tracing Bodies through Liminal Blends during Play-based Inquiry in a Mixed Reality Environment ........ 1431
Danielle Keifert, Noel Enyedy, Joshua Danish, Christine Lee, Maggie Dahn, Lindsay Lindberg

Curiosity Practice: A Powerful New Lever for Fostering Science Engagement ................................. 1433
Leema K. Berland, Rosemary S. Russ, Noah Weeth Feinstein

The Assessment of Digital Reading Skills With Cognitive Diagnose for the Reading Achievement
Test in China ..................................................................................................................................... 1435
Yan Liu, Chuxin Fu, Xiaqing Gu

Translating Theory to Practice: Technology Solutions to Solve Practical Issues for Teaching Reading
Comprehension at the Secondary Level ......................................................................................... 1437
Donna Caccamise, Megan K. Littrell-Baez, John Weatherly

Student Learning of Computational Thinking in A Robotics Curriculum: Transferrable Skills and
Relevant Factors ............................................................................................................................... 1439
Guanhua Chen, Ji Shen

The Challenge of Working With the Future Within STEM Education ............................................. 1441
Giulia Tasquier, Olivia Levrtini, Antti Laherto, Caitlin Wilson, Elina Palmgren

The Role of Instructional Goal Setting for Teaching Computational Thinking in Robotics Classrooms...... 1443
Eben Witherspoon

Concrete Definition of Beneficial Collaborative Dialogues ........................................................... 1445
Michelene T. H. Chi

Concepts Before Coding: The Impact of Classroom Culture and Activity Design on Student Engagement
With Computer Science Concepts ........................................................................................................ 1447
Patrik Lundh, Shuchi Grover, Nicholas Jackiw
Multimodal Reflection: Adolescents Remixing and Sharing their Experiences in an Informal STEM+L Academy

Blaine Smith, Ji Shen, Shiyan Jiang, Guanhua Chen, Marie Hamaoui, Juan Torralba

Eliciting Student Explanations in an Undergraduate Biology Laboratory Course

Anna Strimaitis Grinath, Sherry A. Southerland

Developing a Text-Integration Task for Investigating and Teaching Interdisciplinarity in Science Teams

Simon Knight, Kate Thompson

Making Uncertainty Work: How Youth Manage Uncertainty to Shape Learning Trajectories in a School Makerspace

Colin G. Dixon, Lee M. Martin, Sagit Betser

Interdependence as a Treatment Effect: An Example From Group Awareness Research

Lenka Schnaubert, Daniel Bodemer

ORBIT - Overcoming Breakdowns in Teams With Interactive Tabletops

Patrick Sunnen, Béatrice Arend, Valerie Maquil

Growing Teamwork Competency: A Mixed Methods Study of an Iterative Digital Formative Assessment Approach

Elizabeth Koh, Jennifer Pei-Ling Tan, Helen Hong, Yi Huan Tee

Visualizing Knowledge in the Era of Instructional Software and Gamification: Challenges in Design, Method and Practical Use

Eva-Maria Ternblad, Agneta Gulz

Unpacking Why Student Writing Does Not Match Their Science Inquiry Experimentation in Inq-ITS

Haiying Li, Janice Gobert, Rachel Dickler

Science Literacy in Controversial Contexts: An Epistemic Balancing Act

Aviv J. Sharon, Ayelet Baram-Tsabari

Navigating “Disability” in “Intensive Instruction”: Learner Complexity and Small Environments

Jessica Heather Hunt

Increasing the Use of Formative Feedback: Utilizing Game-Based Principles

Man-Wai Chu, Teresa Anne Fowler

Measuring Awe and Critical Thinking in a Science Museum

Aaron Price, Jana Greenslit, Lauren R. Applebaum, Gloria Alicia Segovia, Chaucey Slagel, Kimberly A. Quinn, Sheila Krogh-Jespersen

A Look at the First Two Years of a 5-Year Longitudinal Study of an OST Program’s Impact on STEM Career Interest

Aaron Price, Angela D. Skeele-Worley, Robert Tai

Talking Past One Another: Looking for Signs of Conversational Mismatch in One 6th Grade Science Classroom

Mon-Lin Ko, Andrew Elby

Uncertainty Management in Science Argumentation

Catherine E. Cullocott, Ying-Chih Chen

The Santa Trap: When Scaffolding Is Not Enough to Challenge Teachers' Pervasive Beliefs

Chandra Hawley Orrill, Rachael Eriksen Brown

Developing Historical Thinking in PBL Class Supported with Synergistic Scaffolding

Haesol Bae, Fangli Xia, Yuxin Chen, Kalani Craig, Cindy Hmelo Silver

Personal Experience and Emotion in Making Sense of Literary Texts
Teresa Sosa, Allison H. Hall
Frictional Patterns in the Design of Games for Learning ................................................................. 1487
Adam Mechtle, Matthew Berland

BioSCANN: A Collaborative Learning Platform That Scaffolds Scientific Inquiry in the Context of
Interrupted Case Studies ..................................................................................................................... 1489
Leslie Schneider, Berri H. Jacque, Jessica Henry

Characterizing Digital Contexts of Collaborative Learning: An Updated Classification Scheme for
Computer-Mediated Communication .................................................................................................. 1491
Alyssa Wise, Trena Paulus

“My Favorite Part is When We Tell the Truth”: Identity and Agency in Middle School Youth’s Climate
Science Digital Storytelling ................................................................................................................ 1493
Elizabeth Walsh, Elizabeth Smullen, Eugene Cordero

Examining the Role of Unpacking 3-Dimensional Teaching and Learning in Museum-Based Professional
Development ....................................................................................................................................... 1495
Gauri A. Vaishampayan, Aaron Price, Kyle Kauffman, Patricia Messersmith, Laura Rico-Beck, Brian
Douglas Gane

The Impact of a Lego Exhibit on Awareness of the Roles and Identities of Engineers ......................... 1497
Gloria Alicia Segovia, Aaron Price, Jana Greenslit, Rabia Ibtasar

Community Science Identity: Becoming Community Scientists in a Museum-based Genetics of Taste
Program .............................................................................................................................................. 1499
Leighanna Hinojosa, Rebecca D. Swanson, Joseph L. Polman

Bridging Multiple Ecologies to Support and Research Learning in Contested Spaces ......................... 1501
Joanna Weidler-Lewis, Cynthia Story Graville, Mary Gould

Learning to Think Computationally: Comparative Outcomes of a Robotics Workshop for Girls ........... 1503
Florence R. Sullivan, Kevin Keith, Ali Soken, Duy Pham

Measuring Maker Mindset: Establishing Content Validity with Card Sorting ....................................... 1505
Jonathan Cohen, Lauren E. Margulieux, Maggie Renken, Shaunna F. Smith, W. Monty Jones

Preparing Students for Learning Mathematics with Adventure Game: Learning Cycle Model of Gaming,
Watching, and Practicing ................................................................................................................... 1507
Hiroki Oura, Ryohet Ikejiri, Kae Nakaya, Ryota Yamamoto, Yuhei Yamauchi

Recounting Counting: Situated Self-Appraisal in Math-Adjacent Discourse ......................................... 1509
Suraj Uttamchandani, Kylie A. Peppler

Beyond Copy Room Collaboration: A Case Study of Online Informal Teacher Professional Learning ........ 1511
Robin Keturah Anderson

Srinjita Bhaduri, Katie Van Horne, Tamara Sumner, Randy Russell, John D. Ristvey Jr.

Towards a Critical Sociocultural Theory of How Teachers Understand Inequity, Power, and Oppression 1515
Grace A. Chen

Learning Scientific Practices Through Participation as a Volunteer Community Scientist ..................... 1517
Rebecca D. Swanson, Leighanna Hinojosa, Joseph L. Polman

Mobility, Diversity, and Openness: Distilling Design Principles for Equitable and Accessible
Makerspaces ........................................................................................................................................ 1519
Kylie A. Peppler, Anna Keune, Justin Whiting

Developing Interdisciplinary Competencies for Science Teaching and Learning: A Teacher-Researcher
Professional Learning Community ......................................................................................................... 1521
Europa Universalis IV the Grandest LAN Party: A Case Study ........................................................................... 1523

Revitalizing Japanese Lesson Study through Shared Tools Embedded in Design-Based Implementation Research .......................................................................................................................... 1525

Improving Elementary Students’ Literacy through Knowledge Building .......................................................... 1527

Investigating the use of Anchoring for Promoting Design Thinking .................................................................... 1529

Learning Loops: Affordances and Challenges of Project Bloks ......................................................................... 1531

Architectures for Learning and Successive Processes of Scaling ........................................................................ 1533

Collaborative, Multi-perspective Historical Writing: The Explanatory Power of a Dialogical Framework ......... 1535

Eye Tracking Students’ Gazes on Feedback in a Digital Assessment Game ......................................................... 1537

Promoting Cognitive Processes of Knowledge Integration ..................................................................................... 1539

Co-Designing Orchestration Support for Social Plane Transitions with Teachers: Balancing Automation and Teacher Autonomy ........................................................................................................... 1541

Exploring How Students Learn Estimation Using a Modelling-based Learning Environment ........................ 1543

Leveraging MOOCs for Blended Learning: Capturing Effective ‘Wrapping’ Strategies With a Learning Design Pattern Language ........................................................................................................... 1545

Leave Some Space to Think: Can Less Guidance Bring More Product? ................................................................. 1547

ERCHI: A Framework for Identifying Observable Signifiers of Student Engagement in Practice Based Learning to Inform the Design of Learning Analytics ........................................................................ 1549

How Learning Outcomes areMeasured in Digital Learning Environments in Higher Education ................. 1551

Rethinking TPACK in the Digital Age: Non-Linear Relationships between Learning by Design, Teachers’ Technology-Related Knowledge and Technology Integration in the Classroom ........................................... 1553

The Difference That Counts: Guiding Knowledge Exchange by Visualizing Levels of Co-Learners’ Knowledge ................................................................................................................................. 1555

“Sorry if I’m Leaving You in the Dust”: Toward Understanding the Importance of Student Goals in Collaborative Problem Solving .......................................................................................... 1557
Mehmet Celepkolu, Kristy Elizabeth Boyer

Biohacking Food: A Case Study of Science Inquiry and Design Reflections About a Synthetic Biology High School Workshop ................................................................. 1559

Justice Toshiba Walker, Mia Shaw, Yasmin Kafai, Debora Lui

How Groups Regulate Their Learning: The Influence of Achievement Goals on Self-, Co- and Shared Regulation Strategies ........................................................................................................................................ 1561

Martin Greisel, Nadine Melzner, Ingo Kollar, Markus Dresel

“Things Are Made to Fail”: Constructive Failures in a Middle School Robotics Curriculum ............... 1563

Andrea Sarah Gomoll, Erin Tolar, Cindy Hmelo Silver, Selma Sabanovic

Using a Video-Based Approach to Develop Pre-Service Science Teachers’ Understanding of How to Teach Nature of Science ................................................................................................................................... 1565

Kennedy Chan, Ka Lok Cheng, Carol Chan, Benny Yung

Videocase Complexity and Preservice Teacher Noticing: Examining the Effects of Cognitive Load .......... 1567

Alison Castro Superfine, John Bragelman

Effects of Expertise on Teachers’ Technology-Supported Teaching Scripts .................................................... 1569

Christina Wekerle, Ingo Kollar

Space for All: Self Construct While Learning in the Digital Age ................................................................. 1571

Tutaleni I. Asino, Penny Thompson, Kathy Essmiller

Supporting SEL Learning in Progressive Design Contexts .................................................................................. 1573

Dhvani Toprani, Marcela Borge, Yu Xia

If You Add Too Much Science, It Gets Boring: Exploring Students’ Conceptual Change in Game Design Iterations ............................................................................................................................................... 1575

Christopher M. Hovey, Camillia Matuk, Talia A. Hurwich

Formative Feedback for Learning. Case Studies of Automated Feedback in Undergraduate Computer Science Education ...................................................................................................................... 1577

Omid Mirmotahari, Crina Damsa, Yngvar Berg

Skyscraper Games: Designing Professional Development for Middle School Teachers to Promote Computational Thinking Using Custom Tools .................................................................................................................................. 1579

Matthew Duvall, Frank J. Lee, Brian Smith

Opening the Door to Algebra: The Role of Fraction Knowledge in Algebra Learning ................................ 1581

Julie L. Booth, Kristie J. Newton, Laura L. Pendergast, Christina Barbieri

How to Enjoy Writing Papers: Supporting Literature-Based Inquiry Learning to Reduce Procrastination and Foster Ownership and Positive Emotions .................................................................................................................................. 1583

Julia Eberle, Tim Schönfeld, Selma Arukovic, Nikol Rummel

Identifying Methods to Induce Productive Confusion for Improving Performance in Physics ................. 1585

Jeremiah Sullins, Katie Console, Rebecca Denton, Clayton Henrichson

Examining Science Identity Development in a Disciplinary Role-taking Multimodal Composing Environment ........................................................................................................................................ 1587

Shiyan Jiang, Ji Shen, Blaine Smith, Kristin Watson Kibler

‘I Think It’s Kind of Made Everybody a Little Closer’: A Virtual Platform Extending Parent Learning as Community of Practice ..................................................................................................................................... 1589

Susan Walker

The Effect of Inhibiting Hand Gestures on Mathematical Reasoning .......................................................... 1591

Candace A. Walkington, Dawn Woods, Mitchell Nathan, Geoffrey Chelule, Min Wang

Measuring Students Epistemic Understanding of, and Beliefs About, Political Media .............................. 1593
Jeremy Stoddard, Jason Chen
Tracking the Flow of Discussion Topics in an Inquiry Science Unit ............................................................... 1595
Catherine Louise Dornfeld Tissenbaum, Sadhana Puntambekar

Supporting the Development of Teacher Candidate Formative Assessment Practice ........................................ 1597
Kate Henson

A Continuum of Knowledge Structures in an Observation-Based Field Geology Setting ....................................... 1599
Lauren Barth-Cohen, Sarah Braden

How Do Kindergarten and Primary School Children Justify Their Decisions on Planning Science Experiments? .......................................................................................................................... 1601
Heidi Haslbeck, Eva-Maria Lankes, Eva S. Fritzsche, Lucia Kohlhauf, Birgit Jana Neuhaus

Transporting Knowledge: A Case Study of Meaning Making on the Pathways of Science Communication ................................................................. 1603
Pryce Davis

Connected Cosplay: Fan Work as Pathways Toward Opportunity ........................................................................ 1605
Sophia Bender, Kylie A. Peppler

An Enhanced Framework for Scale Cognition Leveraging Visual Metaphor Theory and Analogical Reasoning Theory .................................................................................................................. 1607
Cesar Delgado, Matthew Peterson

Visualizing Complex Classrooms Through Real Time Observations ........................................................................ 1609
Joey Huang, Andrea Sarah Gomoll, Erin Tolar, Cindy Hmelo Silver, Selma Sabanovic

Learning With Songo: The African Board Game ........................................................................................................ 1611
Rebecca Yvonne Bayeck

Defining Alternative as More Than At-Risk: Youth Defined Outcomes and Emerging Identities in Alternative Schools ............................................................................................................................ 1613
Gavin Tierney

Neurocognitive Task Analysis: What Changes as Medical Trainees Develop Perceptual Abilities? ........................ 1615
Liam Rourke

GenEvo - An Emergent Systems Microworld for Model-Based Scientific Inquiry in the Context of Genetics and Evolution .................................................................................................................. 1617
Sugat Dabholkar, Gabriella Anton, Uri Wilensky

Re-Engaging Youth: Using Discourse Analysis to Explore Individual Agency and Community Belonging ................................................................................................................................. 1619
Gavin Tierney

Empowering Transformative Agency through Critically Experimenting With Arts in Public Schooling ................. 1621
Raymond Kang

Identifying Shifts in Agency by Analyzing Authority in Classroom Group Discussion ........................................... 1623
Mary Bridget Kustusch, Eleanor C. Sayre, Scott Franklin

A Method for Determining the Extent of Recent Temporal Context in Analyses of Complex, Collaborative Thinking ........................................................................................................................ 1625
Andrew Ruis, Amanda Sieburt-Evestone, Rebecca Pozen, Brendan R. Eagan, David Williamson Shaffer

Rituals, Explorations, and Cultural Resources in the Mathematics Classroom: When Arguing Does Not Help Learning .............................................................................................................................. 1627
Nadav Ehrenfeld, Einat Heyd-Metzuyanim

Investigating Third Grade Students' Collaboration in Project-Based Learning to Inform Curriculum Design ........... 1629
Vanessa Webb, Apriel K. Hodari, Angela Johnson, Rose Nicole Young, Elizabeth Mulvey

Turning to Experience Negative Signs as Operations ................................................................. 1667

Julie Nurnberger-Haag

Authentic to Whom and What? The Role of Authenticity in Project-Based Learning in English Language Arts ................................................................. 1669

Joseph L. Polman, Kristina Stamatis, Alison Boardman, Antero Garcia

When Is It Safe Enough? Considering Diversity and Equity When Brokering Pre-Professional Opportunities to Youth of Color ............................................................... 1671

Rafi Santo, Dixie Ching, Eda Levenson, Geneva White, Mikey Cordero, Kylie A. Peppler, Christoper Hoadley

Projected Worlds: How Informal Digital Learning Organizations Conceptualize Organizing Youth Futures ................................................................. 1673

Rafi Santo, Juan Pablo Sarmiento, June Ahn

Social Reading: Field Study With an In-Home Learning Companion Robot ........................................... 1675

Joseph E. Michaelis, Bilge Mutlu

Engaging Teachers in Discussions Around Temporality Measures from Analytics to Inform Knowledge Building Discourse ................................................................. 1677

Chew Lee Teo, Carol Chan, Dingxuan Ng

Constructing Entities in Scientific Models .................................................................................. 1679

Leslie Atkins Elliott, Lauren Barth-Cohen

Using Example-based PF Conditions to Investigate Preparatory Effects of Problem-solving Prior to Instruction .................................................................................. 1681

Christian Hartmann, Nikol Rummel, Tamara van Gog

Student Perceptions of Object-based Learning With Digitized Museum Materials During Classroom Science Instruction .................................................................................. 1683

Kirsten R. Butcher, Michelle Hudson, Madlyn Runburg

Introducing Bifocal Modeling Framework in Elementary School: Learning Science Using Tangible Modeling Tools .................................................................................. 1685

Tamar Fuhrmann, Engin Bumbacher, Paulo Blikstein

Teaching and Learning Using Virtual Reality: Identifying and Examining Two Design Principles of Effective Instruction .................................................................................. 1687

Britte Haugan Cheng, Cynthia D’Angelo, Sarah Zaner, Matthew Kam, Rhonya A. Hamada

Teacher's Re-design of Virtual Reality Based Curriculum in an Elementary Classroom ......................... 1689

Insook Han

Studying the Interactions between Components of Self Regulated Learning in Open Ended Learning Environments .................................................................................. 1691

Anabil Munshi, Ramkumar Rajendran, Jaclyn Ocumpaugh, Allison Moore, Gautam Biswas

Mentors in the Making: A Case Study of Heterogeneity in Meaning Making at a Public Library Makerspace .................................................................................. 1693

Tesha Sengupta-Irving, Lauren Vogelstein, Corey Brady, Emily Phillips Galloway

Exploring the Impact of Virtual Internships for Democratic and Media Education ............................ 1695

Jeremy Stoddard, Jason Chen

Conceptual Goals While Using a Simulation: Three Different Sources and Learning Outcomes ............... 1697

Robert C. Wallon, Robb Lindgren

Toward the Design of Scaffolds for In-the-World Situated Science Reflections Through Wearables ....... 1699

Insook Han
Sharon Lynn Chu, Brittany Garcia, Beth Nam

Room for Everyone? Identification Processes in Crafting and Math.................................................................1701
Katherine Carr Chapman, Melissa Gresalfi, Amanda M. Bell

Uncovering Students' Ecological Knowledge Resources ....................................................................................1703
Lana Minshew, Kelly Johnson Barber-Lester, Sharon Derry, Janice L. Anderson

Technology to Support Students' Learning Mathematics From Other Students Work .....................................1705
Jeffrey Bush, Brent Milne

Playing Well With Others: An Ethnographic Examination of a Cross-Disciplinary Science-Theatre Collaboration ......................................................................................................................1707
Ariella Flora Suchow

Pedagogically Informed Peer Teaching as a Mechanism for Systematically Maximizing Sociocultural Theories of Learning ........................................................................................................1709
Soren Rosier

Characterizing Chemistry Practices: How Teachers Design and Perform Chemistry Experiments ..............1711
Suna Ryu

Assessment in a Digital Age: Rethinking Multimodal Artefacts in Higher Education ....................................1713
Amani Bell, Jen Scott Curwood, Jen Ross

Engaging With Climate Change as a Socioscientific Issue in an Informal Science Learning Environment ....1715
Kelsey Tayne, Megan K. Littrell-Baez, Erin H. Leckey, Anne U. Gold

Activity Systems Analysis of the Social Practices by Low-achieving Students in a Knowledge-Building Environment Augmented by Reflective Assessment ........................................................................1717
Yuqing Yang, Jan van Aalst, Carol Chan

Criss Crossing Science Domains in Knowledge Building Communities: An Exploratory Study ....................1719
Ahmad Khanlari, Gaoxia Zhu, Stacy Costa, Marlene Scardamalia

“How Do We Pack the World Into Words?” Examining the Collective of Humans and Non-Humans in the Science Classroom ........................................................................................................1721
Donald J. Wink

Investigating Multiple Dimensions of Student Engagement with Embodied Science Learning ...................1723
Megan Humburg, Joshua Danish

Relating Social Network Structure to Uncertainty and Social Interaction in an Engineering Design Challenge .................................................................................................................................1725
Christiana Bruchok, Nicole Bowers, Michelle E. Jordan, Wendy Wakefield, Bernard Ricca

Becoming, Being, and Sometimes Leaving: A Longitudinal Ethnographic Perspective of Climate Scientists’ Participation in Science and Education ...........................................................................1727
Elizabeth Walsh

Shifting Educational Activity Systems: A Cross-Case Analysis of Science Education Reform Efforts in Large Scale Systems ........................................................................................................1729
Deb Morrison, Gina Tesoriero, Philip Bell

Using Web 2.0 Technologies to Facilitate Scaffolding of Student-Led, Collaborative Learning Outside of the Classroom ........................................................................................................................................1731
Stephen M. Rutherford, Amber M. Moorcraft, Sheila L. Amici-Dargan

A Learning Sciences Perspective on the Development of Teachers’ Digital Thinking ....................................1733
Yu-Hui Chang

Making Mathematical Thinking Visible Through Technology ..............................................................................1735
Jeffrey Bush, Brent Milne
Networks in Small-Group and Whole-class Structures in Large Knowledge Building Communities ..........1737
Xueqi Feng, Jan van Aalst, Carol Chan, Yuqin Yang

An Initial Examination of Designed Features to Support Computational Thinking in Commercial Early Childhood Toys ..........................................................1739
Megan M. Marie Hamilton, Jody Clarke-Midura, Jessica F. Shumway, Victor R. Lee

Socialization and Cognitive Apprenticeship in Online Doctoral Programs .............................................1741
Murat Oztok, Kyungmee Lee, Clare Brett

Game Design Literacy as a Problem-Solving Disposition ..............................................................................1743
Beaumie Kim, Reyhanah Bastani

Computational Discourse in a Role-Playing Game Podcast .........................................................................1745
Joshua Gabai, Matthew Berland

Examining Parent-Child Communication and Affect During Tabletop Gameplay in a Children’s Museum:
Implications for Learning ..............................................................................................................................1747
Kristen Missall, Salloni Nanda, Caitlin Coursenth, Benjamin DeVane, Jeremy Dietmeier, Ben J. Miller, Michala Brand

Professional Development of Science Teachers in Underserved Communities: An Initial Report From
the Field .............................................................................................................................................................1749
Tamar Fuhrmann, Cassia Fernandez, Tatiana Hochgreb-Haegele, Paulo Blikstein

“I Have an Opinion about Science I Think Part is True and Part is Not”: Emergent Bilingual/Multilingual
Adolescents “Figuring” Science Learning Through Virtual Labs .................................................................1751
Shakhnoza Kayumova, Suzanne Cardello

Linking Identity Resources Across Roles: Family Science Workshops and Badging ......................................1753
Gavin Tierney, Theresa Horstman, Carrie Tzou

Accessibility, Making and Tactile Robotics: Facilitating Collaborative Learning and Computational
Thinking for Learners with Visual Impairments ..............................................................................................1755
JooYoung Seo, Gabriela Richard

An Analysis of Collective Knowledge Advancement and Emergent Nature of Ideas in Subject-Matter
Learning ............................................................................................................................................................1757
Jun Oshima, Takashi Tsunakawa

Facilitation in Informal Makerspaces ...............................................................................................................1759
Sarah Priscilla Lee, David Bar-El, Kit Martin, Marcelo Worsley

How Do Multilingual Learners Support One Another’s Science Learning and Participation? ..................1761
Mavreen Rose S. Tuvilla, Casey E. Wright, Minjung Ryu, Shannon M. Daniel

The Structures of Embodied Play Activities and Their Impact on Students’ Exploration of the Particulate
Nature of Matter ...............................................................................................................................................1763
Bria Davis, Xintian Tu, Joshua Danish, Noel Enyedy

Assessing the Validity of Peer Feedback in a Sixth Grade Mathematics Class ..............................................1765
Melissa Patchan, Karen Rambo-Hernandez, Brianna Dietz, Kennedy Hathaway

Parents’ Decontextualized Talk During Early Childhood Predicts the Neural Basis of Narrative
Processing in Later Childhood .........................................................................................................................1767
Ozlem Ece Demir-Lira, Salomi S. Assaridou, Susan C. Levine, Susan Goldin-Meadow, Steven L. Small

The Impact of Evidence-Based Professional Development on Classroom Dynamics .................................1769
Gaowei Chen, Carol Chan, Jinjian Yu, Liru Hu, Sherice Clarke, Lauren Resnick

Game-talk: Media-based Mentoring as a Process of Reframing Relationships and Reframing
Perspectives ........................................................................................................................................................1771
Deena Gould, Priyanka Parekh

Using Multiple Perspectives to Study Identity Development in Digital Environments ........................................... 1773
Mirlanda E. Prudent

Examining Productive Discourse and Knowledge Advancement in a Knowledge Building Community .............. 1775
Yuyao Tong, Carol Chan, Jan van Aalst

Crossover Papers
Towards a Framework for Smart Classrooms that Teach Instructors to Teach ................................................ 1779
David Gerritsen, John Zimmerman, Amy Ogan

Augmenting Formative Writing Assessment with Learning Analytics: A Design Abstraction Approach ............ 1783
Simon Knight, Antonette Shibani, Simon Buckingham-Shum

Distributed Representation of Misconceptions ...................................................................................................... 1791
Zachary Pardos, Scott Farrar, John Kolb, Gao Xian Peh, Jong Ha Lee

Envisioning a Learning Analytics for the Learning Sciences .................................................................................. 1799
Alyssa Wise, Yi Cui

Renovating Assessment for the Future: Design-Based Implementation Research for a Learning-in-Class Monitoring System Based on the Learning Sciences ......................................................... 1807
Hajime Shirouzu, Moeji Saito, Shinya Ikubo, Takahiro Nakayama, Kimihiko Hori

Applying Group Communication Analysis to Educational Discourse Interactions at Scale .............................. 1815
Nia Dowell, Oleksandra Poquet, Christopher Brooks

Transhumanism and Education: Embodied Learning in an Era of Altered Bodies ............................................. 1823
Michael Eisenberg

Scaffolding Peer Facilitation in Computer-Supported Problem-Based Learning Environments ......................... 1831
Asmalina Saleh, Cindy Hmelo Silver, Yuxin Chen, Katherine Shanahan, Jonathan Rowe, James Lester

How Gender Cues in Educational Video Impact Participation and Retention ................................................... 1835
Christopher Brooks, Joshua Gardner, Kaifeng Chen

Three Research Directions for Affective Learning Technologies ......................................................................... 1843
Amy L. Baylor

Early Career Workshop
Politicization as Learning: Centring Racialization, Colonialism, and Gender in Learning Sciences Analysis .......... 1849
Joe Curnow

Reasoning About Uncertainty and Efficient Decision-making in Engineering Design ........................................... 1852
Chandan Dasgupta

Studying the Process of Instructional Improvement Through the Lenses of Sense-Making Repertoires and Improvement Practices ................................................................. 1854
Elizabeth B. Dyer

Understanding the Role of Embodied Interaction in Preschool Children’s Learning About Science in Informal Settings ....................................................................................................... 1857
Zayba Ghazali-Mohammed

Identifying Expansive Learning in Democratic Activity: A CHAT/DBR Approach to Community-Based Design Partnerships ..................................................................................... 1860
José W. Meléndez

Young Children’s Inquiry Within and Across Settings ....................................................................................... 1863
Danielle Teodora Keifert
Understanding, Redefining, and Designing for Broadening Participation ........................................................ 1866
Déana Aeolani Scipio
Creativity as a Lens to Frame Teachers’ Use of Games for Learning .............................................................. 1869
Mamta Shah
Defining, Designing, and Documenting Computational Thinking Across K-12 Education ............................. 1873
David Weintrop
Doctoral Consortium
Fabric-based Computing: New Materials for Learning Computer Science ...................................................... 1879
Anna Keune
Supporting Undergraduate Bioscience Learners in Problem-Solving Process Skills Using a Technology-Enhanced Learning Environment ................................................................. 1880
Anurag Deep
Knowledge Places: Embedding Knowledge in the Space of the Classroom .................................................... 1881
Anthony Perritano
Using Multiple Embodied Representations to Support Learners in Making Connections Across Modeling Activities ..................................................................................................................... 1882
Chris Georgen
Why Can’t We All Just Get Along?: Focusing on Socioemotional Climates to Understand Emotions in Collaborative Learning ..................................................................................................................... 1883
Nikki G. Lobczowski
Collaborative Writing in Higher Education: Investigating the Implementation of CSCL Tools and the Role of Prior Individual Experiences and Preferences ................................................................. 1884
Nore De Grez, Bram De Wever
The Effects of the Productive “Visible-Annotation Tool” (P-“VAT”) for Collaborative Knowledge Construction on Higher-Order Interaction and Collaborative Outcome ..................................................................................................................... 1885
Yoonhee Shin
Computation, Constructivism, and Curriculum Design ........................................................................................ 1886
Shayan Doroudi
Exploring the Definition and Measurement of Collaborative Problem Solving ...................................................... 1887
Kristin Stoeffler
Making Energy Easy: Interacting With the Forces Underlying Chemical Bonding Using the ELI-Chem Simulation ..................................................................................................................................................... 1888
Asnat R. Zohar, Sharona T. Levy
Indexes
Author Index ................................................................................................................................................ A1-A11
Keyword Index ............................................................................................................................................ K1-K16
Keynote Presentations
Digital Footprints and Shoes that Don’t Fit

Iris Tabak, Department of Education, Ben-Gurion University of the Negev, Israel, itabak@bgumail.bgu.ac.il

Abstract: The theme for the 2018 International Conference of the Learning Sciences is – "Rethinking learning in the digital age: Making the Learning Sciences count." With this theme, the hosts of ICLS2018 inspire us to look back, aim forward, and examine the present. In furthering this mission of making the learning sciences count, we need to ask: what does it mean to count? What aspects of experience, activity and learning do we want to privilege and support? Do current technologies and their deployment align well with these ideals or might they displace them? In this talk, I will discuss such questions, and suggest research directions that can mitigate tensions between past commitments and future possibilities.

Iris Tabak studies how intra-personal factors (e.g., affect and sense of self) and inter-personal factors (e.g., material and social interactions) mediate complex reasoning. Her Teacher as Partner research examines how to foster learners who feel able and entitled to engage in science. In the past, she conceptualized synergistic scaffolding and developed The Galapagos Finches. Currently, she investigates how non-scientists can use networked resources to make evidence-based health decisions; how epistemic socialization can cultivate a disciplinary stance; and how a focus on multimodal information literacy can make disciplinary socialization relevant to civic participation. Tabak, of Ben-Gurion University of the Negev, holds a B.S.E. in Computer Engineering (University of Michigan), and a Ph.D. in Learning Sciences (Northwestern University). She was a partner in the EC’s FP7 CoReflect project, and the LINKS center (ICORE) on Learning in a Networked Society, was president of the International Society of the Learning Sciences (ISLS), co-editor of the Journal of the Learning Sciences (with Radinsky), and is an ISLS Fellow.
Developing Ethnographic Eyes: Tracing Learning Processes and Practices Across Multiple Levels of Times, Space, and Sociocultural Configurations

Abstract: In the Introductory chapter to the new Handbook of the Learning Sciences, the editors—Frank Fischer, Susan Goldman, Cindy Hmelo-Silver and Peter Reisman frame the learning sciences as an interdisciplinary field that has emerged over the past 25 years as “a nexus of research on how people learn, what might be important for them to learn and why, how we might create contexts in which such learning can occur, and how we can determine what learning has occurred and for whom.” Over the past three decades, interactional ethnography (IE) has developed a parallel program of research. In addressing these goals, the IE research community has developed an interdisciplinary logic-of-inquiry, an epistemology, for studying these questions and other issues of interest to the Learning Sciences today across and within a range of complex educational spaces. In this talk, I will present three telling cases that complement the IE chapter that Susan Bridges and I wrote for the Handbook of the Learning Sciences (see also NAPLeS site for video introductions). These telling cases make transparent the levels of analysis that IE as a logic-of-inquiry supports. Through these telling cases, I make visible the graphic (re)presentations of processes that support empirical analysis of developing collective and individuals-within-the collective learning opportunities that form a foundation for making theoretical inferences about what constitutes as well as counts as learning processes and practices.

These telling cases are based on studies guided by IE in different sites – in a HS physics lab, a process of becoming an historian in a 5th grade bilingual class, and a General Chemistry course at the US Military Academy. These telling cases also make visible how a common epistemological approach (a set of orienting theories), when used to guide studies in different educational sites. In these cases, I provide evidence of how different theoretical perspectives from anthropology, ethnography in education, interactional sociolinguistics, critical discourse analysis, cognitive science (semantic theory) and sociology (ethnomethodology) are critical in tracing opportunities for learning and how these opportunities are formulated and reformulated (or not) by particular actors across times and events. Each case makes visible the social, cultural and discursive processes and practices that shape, and were shaped by members within and across times and sociocultural configuration of events and actors within events. By tracing what and how, when, where, under what conditions and for what purposes, participants propose, recognize, acknowledge and interactionally accomplish particular ways of engaging in developing sociocultural events, I seek to make transparent why overtime analyses of discourse and social actions are critical to uncovering the layers of intertextual relationships that form the basis for constructing warranted accounts of what constitutes evidence of learning in particular educational spaces.

Judith Lee Green is Professor Emerita at the University of California, Santa Barbara in the Gevirtz Graduate School of Education, which she joined in 1990. At the GGSE, she founded the Santa Barbara Classroom Discourse Group with Carol Dixon and the Center for Literacy and Learning in Networking Communities (L²INC) (https://education.ucsb.edu/judith-green). She co-edited three volumes of the Review of Research in Education on Rethinking Learning: What counts as learning and what learning counts (Green & Luke, Volume 30, 2006), What Counts as Knowledge in Educational Settings: Disciplinary Knowledge, Assessment and Curriculum (Kelly, Luke & Green, Volume 32, 2008), and What Counts as Evidence and Equity (Luke, Green & Kelly, Volume 34, 2010). She co-edited (Green, Camilli & Elmore, 2006), the Handbook of Complementary Methods in Education Research. She was co-founder of the Language and Social Processes SIG, American Educational Research Association, and co-editor of the Reading Research Quarterly (1990-1995). She published two edited volumes on ethnographic and discourse research: Green & Wallat (1981), Ethnography and Language in Educational Settings, and Green & Harker (1998), Multiple Perspective Analysis of Classroom Discourse, and two editions of Multidisciplinary Perspectives on Literacy Research (Beach, Green, Kamil & Shanahan, 1993; 2006). She is currently building a new role for interactional ethnographic research -- internal-external ethnographers, with national and international colleagues: Susan Bridges and Monaliza Chian (University of Hong Kong); Liuda Rupsiene, Ingrida Baranauskiene, and Audra Skukauskaite (University of Klaipeda, Lithuania); Maria Lucia Castanheira (UFMG), and Melinda Kalainoff (US Military Academy), and Stephanie Couch (MIT).
Abstract: Bit by bit, a data-intensive substrate for education is being designed, plumbed in and switched on, powered by digital data from an expanding sensor array, data science and artificial intelligence. The configurations of educational institutions, technologies, scientific practices, ethics policies and companies can be usefully framed as the emergence of a new “knowledge infrastructure” (Paul Edwards).

The idea that we may be transitioning into significantly new ways of knowing – about learning and learners – is both exciting and daunting, because new knowledge infrastructures redefine roles and redistribute power, raising many important questions. For instance, assuming that we want to shape this infrastructure, how do we engage with the teams designing the platforms our schools and universities may be using next year? Who owns the data and algorithms, and in what senses can an analytics/AI-powered learning system be ‘accountable’? How do we empower all stakeholders to engage in the design process? Since digital infrastructure fades quickly into the background, how can researchers, educators and learners engage with it mindfully? If we want to work in “Pasteur’s Quadrant” (Donald Stokes), we must go beyond learning analytics that answer research questions, to deliver valued services to frontline educational users; but how are universities accelerating the analytics innovation to infrastructure transition?

Wrestling with these questions, the learning analytics community has evolved since its first international conference in 2011, at the intersection of learning and data science, and an explicit concern with those human factors, at many scales, that make or break the design and adoption of new educational tools. We are forging open source platforms, links with commercial providers, and collaborations with the diverse disciplines that feed into educational data science. In the context of ICLS, our dialogue with the learning sciences must continue to deepen to ensure that together we influence this knowledge infrastructure to advance the interests of all stakeholders, including learners, educators, researchers and leaders.

Speaking from the perspective of leading an institutional analytics innovation centre, I hope that our experiences designing code, competencies and culture for learning analytics sheds helpful light on these questions.

Simon Buckingham Shum is Professor of Learning Informatics at the University of Technology Sydney, which he joined in August 2014 as inaugural director of the Connected Intelligence Centre: https://utscic.edu.au. Prior to this, he was Professor of Learning Informatics and Associate Director (Technology) at the UK Open University’s Knowledge Media Institute. He brings a background in Psychology, Ergonomics and Human-Computer Interaction, and a career-long fascination with making thinking visible using software. He co-founded the Compendium Institute to connect the international community using his team’s Compendium visual hypermedia tool, used widely for Dialogue, Issue and Argument Mapping in both education and business. He co-edited Visualizing Argumentation (2003, with Kirschner & Carr) followed by Knowledge Cartography (2008, 3rd Edition now in prep., with Okada & Sherborne), and wrote Constructing Knowledge Art (2015, with Selvin). He has been active in shaping the field of Learning Analytics since the inaugural LAK 2011 conference, serving as a Program Chair (2012/2018), convening many workshops, and a regular keynote speaker. He co-founded the Society for Learning Analytics Research, serving as a V-P and on the Executive. Homepage: http://Simon.BuckinghamShum.net
Full Papers
Planning to Iterate: Supporting Iterative Practices for Real-world Ill-structured Problem-solving

Daniel G. Rees Lewis, Jamie Gorson, Leesha V. Maliakal, Spencer E. Carlson, Elizabeth M. Gerber, Christopher K. Riesbeck, and Matthew W. Easterday
daniel.rees.lewis@u.northwestern.edu, jgorson@u.northwestern.edu, lmaliakal@u.northwestern.edu, S.C.@u.northwestern.edu, egerber@northwestern.edu, c-riesbeck@northwestern.edu, easterday@northwestern.edu
Northwestern University

Abstract: Solving real-world highly ill-structured problems involves iteration: gathering information, building, testing, and revising products, experiments, and theories. However, we do not know how to create learning environments to teach iteration for highly ill-structured problems. How might we help student teams effectively iterate for highly ill-structured design problems? In this design-based research study we built on learning sciences research to implement Planning to Iterate—a weekly planning session in which teams create problem and planning representations. The study took place in a 6-week extracurricular undergraduate design program with five undergraduate project teams working on highly ill-structured problems. To understand team iterative practices, we analyzed videos of teams’ weekly planning sessions, and teams’ artifacts. Students significantly increased iterative practices, but infrequently integrated the practices together, suggesting re-design with additional coaching.

Keywords: Iteration, problem-solving, planning tools, design research, project-based learning

Introduction
Iteration—building, testing, and refining products, experiments, and theories—is an essential practice to understand problems and create solutions (Schön 1993). Scientists iterate on experiments to create new knowledge, engineers iterate on products for technical and social outcomes, and policy analysts iterate upon policy briefs to create arguments for how policies can achieve desired public outcomes.

Despite the importance of iteration across disciplines, novices struggle to iterate and so waste valuable time pursuing ineffective solutions. Consider the following undergraduate design team working on a highly ill-structured problem. The team partnered with a chain of non-profit stores that sell recycled building materials to reduce CO2 emissions. In the first week, the team identified a problem: most customers left shortly after entering stores without buying materials. By week 2, the team proposed signage to help customers better navigate the store and spent the next 4 weeks designing the signage and an online platform for distributing the signs. In week 5, the team deployed the signage in a local store and found that the signs didn’t change customer behavior because they added to the clutter and went unnoticed. With only one week remaining, the team had no time to shift to a new solution and the problem went unsolved. The team could have tested and learned from a lower fidelity version of the signage earlier to uncover this unknown. The team’s choice to avoid iterating is common; this team was the first of the four teams in their class to test solutions. So why was iterating so challenging?

Iteration was challenging because the teams were taking on real-world highly ill-structured problems—problems with unclear parameters that require defining the problem before it can be solved (Jonassen & Hung, 2015; Simon 1973). The nature of highly ill-structured problems mean (a) goals for iterating are unclear and shift, (b) few tests can be quickly conducted in the classroom, and (c) educators cannot tailor problem-specific scaffolds as they do not know the solution. In this paper, we focus on iteration in design problems: design problems are a broad category of highly ill-structured problems in fields such as education, science, and engineering in which problem-solvers change “existing situations into preferred ones” (Simon 1996, p. 111). Conducting effective iterations for highly ill-structured problems is so complex that some argue it is too challenging for K-16 students (Crismond & Adams, 2012). In this study, we draw on seminal learning sciences (LS) research that support iteration (e.g. Brown & Campione, 1994; Kolodner et al., 2003; Hmelo-Silver 2004) to ask, how can we help student teams to iterate as they take on highly ill-structured problems?

Background
The shifting nature of highly ill-structured problems makes it difficult for problem solvers to set iteration goals, conduct tests, and for educators to scaffold. As Bardach describes “you will probably keep changing your problem definition, as well as your menu of alternatives, your set of evaluative criteria, and your sense of what evidence
bears on the problem. With each successive iteration you will become a bit more confident that you are on the right track, that you are focusing on the right question.” (Bardach & Patashnik, 2015, p. xv). Unsurprisingly, the more ill-structured the problem, the more experts iterate (Jin & Chusilp, 2006).

Real world, highly ill-structured problems: involve stakeholders outside of the classroom who desire a solution; require solutions unknown to students, teachers, and stakeholders (Simon 1973); and are embedded in real world complex systems. While challenging for students and teachers, highly ill-structured problems provide authentic learning opportunities in the practices and modes of thinking of many disciplines. We present an example project to further define and illustrate highly ill-structured problems, and how they pose difficulties for teaching and learning iterative practices. A student design team partnered with a disability advocacy group, and worked with airline staff, baggage handlers, and wheelchair users to reduce damage to wheelchairs transported in the hold of passenger aircrafts. This problem is ill-structured along several dimensions (Jonassen & Hung, 2015). First, this challenge has high intransparency: there are many important unknowns (e.g. airline policy, baggage handler training); the instructor and students do not know these unknowns or the form the solution might take. Second, this challenge has many legitimate competing alternatives: there are many possible solutions. At the outset, the solution could feasibly include graphic stickers to guide baggage handlers, lifting devices for baggage handlers, airline policy changes, straps for securing wheelchair in the hold, and training. High intransparency and many legitimate competing alternative solutions makes it challenging to know what to iterate on, so the goals and criteria for the iteration are unclear, and educators cannot create solution-specific scaffolds. Third, this challenge has high heterogeneity of interpretations: the multiple stakeholders have different constraints and definitions of success that need to be considered and uncovered (e.g. airlines, advocacy groups, baggage handlers). It is difficult to rapidly iterate as stakeholders are outside of the classroom. Due to the combination of high intransparency, legitimate competing alternatives, and heterogeneity of interpretations, iterating is simultaneously more important and more difficult when solving highly ill-structured problems.

Expert iterative practices to inform learning environments

Schön’s theory of reflective practice (1993) emphasizes experts’ reflection while iterating to inform both their conception of the problem and solution. Learning scientists have built on Schön to operationalize core expert iterative practices and establish expert-novice differences when iterating to structure (Simon 1973) highly ill-structured problems. First, expert designers dedicate time to representing and revising their understanding of the problem (Atman et al., 2007; Jain & Sobek, 2006). In comparison, students dedicate less time to revising their understanding (Atman et al., 2007). Second, experts prioritize, which involves identifying the unknowns most needed to understand the problem, and then iterating to uncover these unknowns first (Adams et al., 2003; Guindon 1990). Representing and prioritizing allows experts to overcome the ambiguity caused by the plethora of options in highly ill-structured problems, by identifying and selecting the important unknowns to uncover that set the goals of the iteration. In comparison students do not uncover unknowns because they iterate without a specific goal (Crismond & Adams, 2012), or respond with inaction in the face of problem ambiguity (Lande & Leifer, 2010). Third, Adams et al. (2003) found experts conduct coupled iteration. Coupled iterations involve learning about both the problem and the solution through creating, testing, and revising prototypes. Fourth, experts conduct cascaded iterations. Cascaded iterations involve rapidly cycling through a problem solving process, including developing solutions, gathering information about the problem or solution viability, and revising solutions (Atman et al., 2007). In comparison to experts, students often do not view rapid iterations as a useful practice (Crismond & Adams, 2012), and so gather less information, and develop and test fewer solutions (Atman et al., 2007). Coupled and cascaded iterative practices allow experts to understand highly ill-structured problems by quickly uncovering unknowns, discarding non-viable solutions, and developing viable solutions.

Teaching iteration

Despite the importance of learning iteration, many educators plan iterations for students through common waterfall (test solutions at the end of the project) and stage-gated (e.g. test solutions by a given week) rather than teaching students to plan iterations (Crismond & Adams, 2012). We build on seminal LS approaches for teaching iteration for moderately ill-structured challenges, be that testing medical diagnostic theories (Hmelo-Silver 2004), scientific theories with peers (Brown & Campione, 1994), or model car designs and physics theories (Kolodner et al., 2003). However, these effective approaches are not designed for highly ill-structured problems, as they (a) provide students with clear goals for iterating (e.g. model car performance), (b) can be quickly tested in the classroom, and (c) allow educators to tailor problem-specific scaffolds as educators know the problem solution (e.g. cases relevant to the solution). Areas such as Maker education also engage students in iterative making (Kafai, Fields, & Searle, 2014), by lowering the barrier to engagement. Extending seminal LS approaches to highly ill-structured challenges, where students solve real-world problems for others, is a demanding task;
Jonassen and Hung noted “very ill-structured and complex problems, like design problems, may be too difficult to learn in a PBL [problem-based learning] setting” (2015, p.22).

Planning to Iterate: Design argument for supporting student-led iteration

How might we help student teams to practice cascading and coupled iterations as they take on highly ill-structured problems? Our Planning to Iterate design argument in this design-based research study (Easterday, Rees Lewis, & Gerber, 2017) proposes that we can support iteration by (a) facilitating discussions and prompting teams to conduct a planning process of representing the problem, prioritizing unknowns, and making iteration plans; (b) using two templates, the design canvas and iteration plan; (c) providing and prompting use of guiding questions that drive students to consider common pitfalls, and examples (Table 1). The Planning to Iterate process we created supports expert iterative practices of representing and revising problem understanding (Atman et al., 2007; Jain & Sobek, 2006) and identifying and prioritizing the unknowns (Guindon 1990) that guide the goals of a coupled and cascaded iteration (Figure 1; Table 1). The design canvas and iteration plan templates help teams consider the important aspects and complexity of the problem (problematizing, Reiser 2004). The supports include the examples and guided questions to help the teams conduct the process and use the templates (Kolodner et al., 2003). Facilitators lead discussions of cases and prompt the process and use of tools, but each team manages their own planning (Brown & Campione, 1994).

Table 1: Features of the Planning to Iterate design argument for supporting expert iterative practices

<table>
<thead>
<tr>
<th>1. Iterative practices</th>
<th>2. Learning environment features</th>
<th>3. Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedicate time to updating what they know about the problem, solution, and which aspects of the problem are important unknowns. (Atman et al., 2007; Jain &amp; Sobek, 2006).</td>
<td>A. Weekly 1-2hr facilitated team planning sessions where teams revise the design canvas, identify important unknowns, &amp; plan iteration</td>
<td>Regular prompted routines help students to learn practices (Brown &amp; Campione, 1994; Puntambekar &amp; Kolodner, 2005)</td>
</tr>
<tr>
<td>Prioritize most important unknowns in the project to define the goals for the iteration (Guindon 1990).</td>
<td>B. Teams update the design canvas problem representation template (figure 1, figure 2b), each session, including noting unknowns</td>
<td>Surfacing complexity (Problematizing; (Reiser 2004); surface project aspects for collaboration (project board, Kolodner et al., 2003)</td>
</tr>
<tr>
<td>Plan to conduct coupled and cascaded iterations. (Adams et al., 2003; Atman et al., 2007; Guindon 1990)</td>
<td>C. Scaffolds representing problem: guiding questions, example design questions &amp; facilitated discussion on A4 sheets</td>
<td>Cases &amp; guiding questions for problem-solving decisions (Hmelo-Silver 2004; Kolodner et al., 2003; Puntambekar &amp; Kolodner, 2005).</td>
</tr>
<tr>
<td></td>
<td>D. Prioritization facilitation: Teams prompted to note which sections of design canvas are the most important unknowns using cases and guiding questions on A4 sheets</td>
<td>Cases of team prioritization and guiding questions (Kolodner et al., 2003; Puntambekar &amp; Kolodner, 2005)</td>
</tr>
<tr>
<td></td>
<td>E. Teams use the iteration plan, a planning representation template, to articulate goals, criteria, and tasks based on prioritization of unknowns (figure 1)</td>
<td>Surfacing complexity (Problematizing, Reiser 2004) surface project element for collaboration (project board, Kolodner 2003)</td>
</tr>
<tr>
<td></td>
<td>F. Example iteration plans guiding questions, &amp; example cases and scaffolds for coupled &amp; cascaded iterations</td>
<td>Cases &amp; guiding questions for problem-solving decisions (Hmelo-Silver 2004; Kolodner et al., 2003; Puntambekar &amp; Kolodner, 2005)</td>
</tr>
</tbody>
</table>

Methods

In this design-based research study (Design Based Research Collective 2003) we implemented Planning to Iterate (figure 1) in a 6-week, full-time (9am-5pm) extracurricular (student-led, no grades/credits) university human-centered design program where 4-5 member student teams designed novel products and services to solve a real-world, highly ill-structured problems. Students included 21 undergraduates (4 first-, 12 second-, 5 third-years; 57% female) ages 18-22 years old at a US university. Students majored or double-majored in engineering (15), sciences (3), social sciences (3), art (3) and journalism (3), and had between 0-2 years of design training.

Before the program, an undergraduate student organizer worked with local partner organizations to scope pro-social problems that the partner organizations needed solving, including: (a) reducing teen depression with a medical research center, (b) increasing mental health first responder support for youth with a medical research center, (c) improving airport accessibility for autistic travellers with a disability advocacy group, (d) reducing air travel related wheelchair breakages with a disability advocacy group, and (e) reducing dementia related stigma with a dementia advocacy group. The undergraduate student organizer also coordinated 2-4 hours of methods workshops (e.g. user interviewers) each week led by design educators.
Planning to Iterate is designed to support teams to revise their problem representation and identify important unknowns that give them goals for their iteration. Simplified example from week 2-3 cycle.

Two of the authors facilitated 1-2 hour weekly planning sessions in weeks 1-5, in which teams update their findings (forthwith ‘learnings’) from the previous cycle on the iteration plan and transfer it to the design canvas (20-35 minutes), engage in a facilitated discussion of cases as a whole class and further revise their canvas with guided question sheets (20-30 minutes), identify unknowns (‘risks’) on the canvas (10-20 minutes), prioritize their most important unknowns/risks (5-10 minutes), and plan their next cycle on the iteration plan (15-25 minutes). Each team had a design canvas and iteration plan fixed to the wall (Figure 2). Teams work on their projects for the next week or ‘cycle’, the result of which they record at the start of the planning session the next week. Updating the design canvas is both the end of one cycle, and the starting point of the next cycle.

Data collection and analysis
The first analysis examined artifacts and video extracts for differences in coupled and cascaded iterations between the Planning to Iterate implementation and the previous year of the program (an earlier iteration of our DBR project). After finding a significant increase in iterative practices between the Planning to Iterate and previous year, we conducted a second analysis of video data of two planning sessions to examine how the implementation supported iteration. We collected videos of planning sessions of all teams (one camera per-team, 23 sessions, 34 hours), and photos of the iteration plan and design canvas (pre- and post planning session).

Analysis 1: Comparing iterative practices between Planning to Iterate and benchmark
We first examined weekly planning sessions to see if teams conducted coupled and cascaded iterations. In each session, teams summarized learning from the previous week and planned design activities for the current week. We collected data on 4 cycles from 5 teams for a total of 18 sessions (because 2 teams missed 1 session each). For each session we analyzed: (a) team learnings captured on the iteration plan (photos), (b) team learnings during the planning session (5-15 minute videos), and (c) the design canvases before and after the session.

To see if teams conducted coupled iteration—gathering information on problems and solution by building and testing to inform revisions—we developed three sub-codes: (a) problem learning to capture team reports of learning about aspects of the problem through testing prototypes, (b) solution learning to capture reports of learning about the solution through testing prototypes, and (c) solution revision to capture reports of plans to change the solution based on reported learning. We coded a cycle as involving coupled iteration if problem learning (code a) and solution learning (code b), informed solution revisions (code c).

To examine cascading iterations—the expert practice of rapidly cycling through a problem solving process—we coded iteration plans, design canvases, and discussion of learnings. Because students were practicing human-centered design, we used user interviews as a measure of gathering information, prototypes as a measure of the number of solutions created, and user tests (with a prototype) as a measure of testing solutions. For each cycle we counted prototypes for each generated or altered prototype. We counted user interviews for each unique user teams interviewed about the problem in the given cycle. We counted user tests for each test of a prototype with one user; if teams tested three prototypes with two users, we counted six user tests.
To provide a comparison of coupled and cascaded iterative practices, we analyzed data from the previous year’s program, which followed a stage-gated approach; teams presented prototypes in weeks 4 and 6. Teams had 80-120 minute weekly coaching sessions, in which teams discussed and recorded their previous learnings and goals for uncovering unknowns. While teams were encouraged by their coach to iterate, the previous program differed from the current study as teams did not receive scaffolding on iteration, and did not engage in a Planning to Iterate process. The program otherwise had the same features, including multidisciplinary teams of 4-5, working full-time on real-world highly ill-structured problems with a client, and the same methods workshops. This study gave us parallel data to a different condition to the current study, namely: the reported learnings on the week's activities, and corresponding audio recording of teams reported learnings. We conducted the same analysis on the previous year’s data as described above.

Analysis 1: Comparison of practices between Planning to Iterate and benchmark
The teams in the Planning to Iterate implementation showed more iterative practices compared to teams in the previous year. Analysis of student artifacts and video or audio data showed that teams were more likely to demonstrate coupled iteration through discussing and recording learnings about both the problem and the solution through creating and testing prototype solutions, and then suggesting the revision of future solutions (Table 2, column 1). In the Planning to Iterate implementation, teams demonstrated coupled iteration in 66% of cycles (12/18); four of the six cycles that did not involve coupled iteration were in the first week. In comparison, teams in the previous year demonstrated coupled iteration in 31% of cycles (5/16). In the Planning to Iterate implementation, all teams conducted coupled iterations in the week 2-3 and 4-5 cycles, whereas in the previous year’s program teams did not start conducting coupled iterations until the week 3-4 cycle (3rd cycle).

Consider an example of a coupled iteration from a team working on teen depression and suicide in the Planning to Iterate intervention. In week 2 the team tested three prototypes of an “emotions planner” designed to foster conversations between middle schoolers and parents building on work in child psychologists on “building emotional vocabulary” (design canvas). On the iteration plan the teams reported seeking information from “middle school students”, “parents”, and a “child psychologist”. The teams wrote on the iteration plan “we got feedback that kids would only fill out the planner if someone enforced it”, and discussed how this information emerged from showing their prototypes to a student and parent (video data). On the iteration plan the team had a post-it note, which read “look into ‘digital’ planner and how it could be implemented”. Teams discussed creating email prompts, and getting teachers or parents to initiate conversations (video data). Finally, on the design canvas the team added a note reading: “weekly mental health related prompts aim to increase communication b/w students + parents → form habit”. This exemplifies coupled iteration: the team discussed a) how problem understanding changed (students don’t start conversations about emotions), b) current design (the design needed to prompt students), and c) how to change prototypes (add human/email prompts).

Findings and discussion
Analysis 2: Examining use of scaffolds
Analysis 1 showed a significant increase in coupled and cascaded iterations (see findings Table 2); as a result, our goal was to conduct a deductive video analysis (Derry et al., 2010) to understand if the design argument was functioning as we hypothesized. Analysis 2 examined video of two planning session from two Planning to Iterate teams: a team working on reducing wheelchair breakages in air travel (113-minutes), and a team working on improving service workers support of dementia sufferers (117-minutes). The planning sessions were taken from week 3 when teams were familiar with the planning sessions. We selected these two sessions because one team (wheelchair breakages) did conduct a cascaded iteration that cycle and one team (dementia stigma) did not. We selected these teams as initial analysis suggested they engaged the least and most in iterative practices. This allowed us to examine ways in which the design succeeded and failed to support student iteration.

We first identified macro-level codes, identifying sections of teams’ planning sessions by sustained discussion and activity a given topic (Derry et al., 2010). We identified 30 (wheelchair team) and 34 (dementia team) sections (30-seconds to 11 minutes). Second, to see if the team engaged in iterative practices we identified which macro-level codes included events in which teams engaged in the iterative practices of: (a) revising aspects of problem representation, (b) identifying unknowns, (c) prioritizing unknowns as goals of the iteration, and planning (d) coupled iterations and (e) cascading iterations. Third, we examined whether teams connected these practices as predicted by the design argument: did teams revise problem understanding to inform their prioritization of unknowns, and then goals of coupled and cascaded iteration. We coded each significant event where team members referenced (verbal/gestures) previous discussions and artifacts (e.g. design canvas).
Table 2: Planning to Iterate teams showed more iterative practices compared to the previous year’s program

<table>
<thead>
<tr>
<th>Condition</th>
<th>Cycle</th>
<th>Percentage of teams conducting coupled iterations each cycle (total/no. of teams)</th>
<th>Mean user interviews of all teams each cycle (standard deviation)</th>
<th>Mean prototypes of all teams built by the teams each cycle (standard deviation)</th>
<th>Mean number of stakeholders tested each cycle (standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous program</td>
<td>Week 1-2</td>
<td>90% (2/2)</td>
<td>6.5 (±2.2)</td>
<td>18 (±9.6)</td>
<td>5 (±2.6)</td>
</tr>
<tr>
<td></td>
<td>Week 2-3</td>
<td>80% (1/2)</td>
<td>6.3 (±2.0)</td>
<td>16 (±8.4)</td>
<td>4 (±1.8)</td>
</tr>
<tr>
<td></td>
<td>Week 3-4</td>
<td>80% (1/2)</td>
<td>6.2 (±2.0)</td>
<td>18 (±9.6)</td>
<td>6 (±2.6)</td>
</tr>
<tr>
<td></td>
<td>Week 4-5</td>
<td>80% (1/2)</td>
<td>6.3 (±2.0)</td>
<td>17 (±9.2)</td>
<td>5 (±2.6)</td>
</tr>
<tr>
<td>Planning to Iterate</td>
<td>Week 1-2</td>
<td>20% (1/2)</td>
<td>3.5 (±0.7)</td>
<td>4 (±0.6)</td>
<td>3 (±0.5)</td>
</tr>
<tr>
<td></td>
<td>Week 2-3</td>
<td>20% (1/2)</td>
<td>3.4 (±0.6)</td>
<td>4 (±0.6)</td>
<td>3 (±0.5)</td>
</tr>
<tr>
<td></td>
<td>Week 3-4</td>
<td>20% (1/2)</td>
<td>3.5 (±0.7)</td>
<td>4 (±0.6)</td>
<td>3 (±0.5)</td>
</tr>
<tr>
<td></td>
<td>Week 4-5</td>
<td>20% (1/2)</td>
<td>3.4 (±0.6)</td>
<td>4 (±0.6)</td>
<td>3 (±0.5)</td>
</tr>
</tbody>
</table>

In the current Planning to Iterate implementation teams conducted far more activity suggesting cascading iterations than in the previous year’s program (Table 2, columns 2-4). As noted above, cascaded iterations involve rapidly cycling through problem solving activities. Based on analysis of video or audio recordings of team discussions, and team’s design canvases and iteration plans, in each cycle teams conducted more user interviews, created more prototypes, and conducted more user tests. In the Planning to Iterate implementation, the mean of every single measure was higher than the previous year’s program in every single corresponding cycle (Table 2, columns 2-4). What might explain these differences?

Analysis 2: Video analysis of two planning sessions to examine scaffold use

To understand why iterative practices increased, we analyzed video data of two teams in Planning to Iterate sessions in week 3 (end of week 2-3 cycle/start of week 3-4 cycle): The wheelchair team (WT) working on wheelchair breakages on passenger planes, and the dementia team (DT) working on helping service workers support dementia sufferers. In both planning sessions teams engaged in the process. Teams discussed and recorded learnings from the previous cycle on the design canvas (instances: WT=12, DT=14). Teams revised their design canvases using the guided questions and examples (verbal/gestural references: WT=7, DT=6). Teams also noted unknowns using guided questions (WT=6, DT=3). Teams set and recorded goals based on the prioritized unknowns in the design canvas (WT=2, DT=1). The teams verbally referenced the examples when planning on the iteration plan, and they planned to conduct coupled and cascaded iterations for the next cycle.

Analysis showed both effectiveness and fragility of Planning to Iterate in supporting iterative practices. The following extract exemplifies WT noting, and nearly missing, the unknown of baggage handler needs. The facilitator has just led a discussion (Table 1). Matt is standing at the design canvas, holding a pen and post-it notes, Jung, Sari, and Aaron are sat at a table facing the canvas. Aaron is holding the guiding questions sheet:

1. Matt: Problem 3. Haven’t said the users needs slash do we know it well? (guiding question)
2. Aaron: I would say so. (Sari: nods)
3. Jung: We know it. We’ve had enough interviews, for now.
4. Matt: Have we added the most current… (…evidence? "=next guiding question on sheet).
5. Jung: …well we don’t know the baggage handlers, needs yet. That’s the one we don’t know.
6. Matt: Oh, that’s true (adds a blank post-it note to user section of design canvas) (…off topic talk…)

Figure 2: 2a. (left) Matt writes on the post-it notes as Aaron holds sheet with guiding questions. 2b. (right) Wheelchair breakages team design canvas and iteration plan (end of the week 3 planning session).
7. Jung: And a so a baggage handler needs…
8. Aaron: A process of work.
9. Matt: (writing on post-it) load…chair… and bags…quickly…lift…heavy…chair…load…onto…ramp…load…get…through…hold. (turns to team). Yeah?
10. Sari: Cool.

In this extract, Jung notes they do not know baggage handlers’ (line 6), after Matt reads aloud a question. In lines 2-4 the all team give verbal or gestural indications they know the user needs. In line 4 Matt starts to read the next question, and Jung interrupts in line 5 to say they don’t know baggage handler needs. The team then articulates and adds baggage handler needs to the canvas. Thirty seconds after this extract Jung says: “Wait, can we actually put risk on the baggage handler, because we actually haven’t spoken to actual baggage handlers”, to which Matt responded “yes we can”, and adds an ‘important unknown’ sign to user needs section. The baggage handler need was referenced four more times over the session in relation to: (1) getting access to baggage handlers (27th minute), (2) wheelchair breakages (38nd minute), (3) solution description (52nd minute), and (4) goals (85th minute). At the end of the planning session Matt pointed at a note on the iteration plan reading build 5 more prototypes, communicate our ideas with baggage handlers and said “this is a crucial thing for us to get done”. Over the next two cycles WT conducted 18 tests with baggage handlers. WT engaged in target practices, but also nearly missed an important unknown with only Jung noting they did not know baggage handler needs.

Video analysis highlighted weakness in supporting teams connect practices. DT raised two and WT raised one important unknown that were then never reference again. For example, in reference to their solution of a training for service workers, DT noted they didn’t know whether organizations (e.g. grocery stores) would support service worker training. A team member, Kacee, noted “we haven’t shown these (pointed at design solutions on design canvas) to any potential partners...so we don’t know if they’ll implement our design”, to which Beth replied “yeah, that’s bad.” The team discussed the issue for 4 minutes, recorded the issue on the canvas, including adding the ‘important unknown’ sign, and did not mention the unknown again in the session.

Discussion

Taken together, the findings suggest Planning to Iterate supports student iterative practices in highly ill-structured problems—practices highly challenging for students enact and for educators to support (Crismond & Adams, 2012). The Planning to Iterate implementation saw more instances of coupled iteration—teams reporting on how their activities informed both problem understanding and solutions. It also saw more cascaded iterative practices—teams conducted more interviews, built more prototypes, and tested more prototypes.

Planning to Iterate might have only changed the distribution but not the aggregate iterative practices relative to a stage-gated approach. That is, we might have seen the same aggregate amount of user research, building and testing, with the Planning to Iterate seeing more balanced activity each week, and the other approaches seeing spikes of activity before deadlines. However, relative to the previous year, in the Planning to Iterate intervention teams conducted more of each activity each week. Our findings suggest Planning to Iterate helped teams surface more unknowns, motivating information gathering. Analysis 2 suggests teams generated multiple goals from unknowns. That both coupled and cascaded iteration increased from the benchmark suggests the design argument can support iteration for highly ill-structured problems.

While Analysis 2 supports aspects of the design argument, fails to help teams make connections between practices. The design argument functioned as hypothesized in that video analysis showed teams followed the Planning to Iterate processes. Furthermore, teams demonstrated iterative practices, such as noting and prioritizing unknowns (Atman et al., 2007; Guindon 1990). That teams articulated and recorded unknowns, and then planned activities to uncover unknowns (Atman et al., 2007), suggests that Planning to Iterate guided teams to set goals for iteration in the face of problems ambiguity (Jonassen & Hung, 2015). For example, WT noted baggage handler needs as an important unknown, and planned to reduce this unknown. However, selecting identified unknowns when planning goals appears to be fragile. For example, DT did not reference that they were unsure if organizations would want their solutions after they raised the unknown. Consequently, DT did not explore this important issue. Similarly, WT might not have noted that they did not understand the baggage handlers needs—an unknown that eventually set the goal for their next two cycles.

These findings also suggest re-designs. Teams can miss important unknowns, so we might add team questioning routines to check peer thinking (Brown & Campione, 1994). Furthermore, problem-based learning (PBL) style coaching (tutoring) could help teams connect the practices (Hmelo-Silver 2004). Such work would answer Jonassen and Hung’s (2015) call to expand PBL to highly ill-structured problems.
Conclusion
This exploratory design-based research study tested Planning to Iterate, a design for supporting teams iterative practices in highly ill-structured problem solving. Findings suggest that Planning to Iterate supports iteration: teams conducted more coupled and cascaded iterations, and the scaffolds supported iterative practices. This work suggests design principles for supporting iterating when solving highly ill-structured problems: representing problem state and important unknowns to guide a plan for iterating. Problem and plan representations are scaffolded through templates, example representations and guided questions. However, fragile processes suggest adding peer support (Brown & Campione, 1994) and PBL coaching (tutoring, Hmelo-Silver 2004). While this work took place in design, highly ill-structured problems occur across domains, so this work could apply to many domains, such as policy analysis and science inquiry. Understanding how to teach students to iterate is critical if we want to prepare them to solve real-world problems that impact our daily lives.

References

Acknowledgements
We thank Delta Lab, Haoqi Zhang, Julie Hui, and Bruce Sherin for their feedback. We thank Alex Sher for help with implementation. This work was funded by the National Science Foundation (US) grants IIS-1530833 and IIS-1320693.
Learning Design Through Science vs. Science Through Design

Elizabeth McBride, Jonathan Vitale, and Marcia C. Linn
bethmcbride@berkeley.edu, jonvitale@berkeley.edu, mclinn@berkeley.edu
University of California, Berkeley

Abstract: This research investigates two ways of framing design projects and their impacts on learning. The study explores the benefits of learning science concepts before or during a design project. Based on the NGSS science and engineering practices, in an engineering condition, students learn the necessary science concepts during a design project. In a science condition, students learn the science concepts first, then apply them during a design project. The study explores the benefits of each approach to inform instructional design. We use the knowledge integration framework to develop curriculum and assessment items, including an interactive computer model of a solar oven. Using three types of pre/posttest assessment items, we found students in both conditions gained insights on science and engineering design items; students in the engineering condition outperformed the science condition on a science-design integration item and conducted more trials during the design process while using an interactive computer model.

Introduction
Engineering projects are becoming more common in K-12 schools, but while it is often claimed that engineering projects improve student achievement in mathematics and science, research on this topic has shown that many projects do not live up to the claim (Teacher Advisory Council, 2009). While engineering projects may generate more student interest and engagement (Hmelo et al., 2000; Cantrell et al., 2006) than typical science curricula, they often fall short on developing science concepts. Ideally, undertaking a science project should be motivating, while also helping students to understand the interplay between science concepts (like energy transformation) and engineering design decisions. However, the framing of goals can impact what aspects of the project are emphasized. In projects framed around science goals, students learn the science concepts and then do a design project to apply those concepts (science condition). In projects framed around engineering goals, students learn the science concepts during the process of completing a design project (engineering condition). We investigate ways these two goal frameworks impacts student learning.

Often, in science the goal is to develop knowledge, while in engineering the goal is to develop a solution (Lewis, 2006; Purzer et al., 2015). In addition, we use the Next Generation Science Standards (NGSS) focus on science and engineering practices, specifically the practice of “constructing explanations (for science) and designing solutions (for engineering)” (NGSS Lead States, 2013) to inform our conditions. This study compares versions of a solar ovens unit that loosely use one or the other goal frames and present a focus on either constructing explanations or designing solutions, while keeping the overall content of the curriculum the same.

We use the knowledge integration framework (Linn & Eylon, 2011) to guide the development of the curriculum and this study. This framework focuses on building a coherent understanding of concepts, and has proven useful for design of instruction featuring dynamic visualizations (Ryoo & Linn, 2012) and engineering design (Chiu et al., 2013; McElhaney & Linn, 2011). The framework emphasizes linking of ideas by eliciting all the ideas students think are important and engaging them in testing and refining their ideas. When students build a physical artifact, as in many engineering projects, they can only test a few of their ideas due to time and material constraints. Features in this curriculum, like using interactive computer models, allow students to explore many more ideas, thereby facilitating knowledge integration.

Though engineering projects are potentially motivating, when students build a physical model they often neglect the scientific basis for their decisions (Crismond, 2001), instead focusing on aesthetic and otherwise superficial details of construction. Tools like interactive computer models can help students connect science principles and design decisions by making mechanisms such as energy transformation visible (Snir, Smith, & Grosslight, 1993; Wilensky & Reisman, 2006). The combination of computer models and hands-on activities in design activities allows students to test many designs while also visualizing how energy transformation takes place in their designs.

In addition to providing science content knowledge, design projects utilizing computer models provide students with an opportunity to explore authentic practices of scientists and engineers. The NGSS envision that instruction would combine practices including modeling, data, analysis, computational thinking, and design to enable students to integrate their scientific and engineering ideas (NGSS Lead States, 2013). The solar ovens
curriculum used in this research familiarizes students with the way energy transforms from solar radiation to heat (MS-PS3-3) by using a hands-on project and interactive models, emphasizing the modeling aspect of the science and engineering practices of the NGSS as well as the standards associated with energy (NGSS Lead States, 2013). This curriculum draws on all eight of the science and engineering practices in the NGSS, focusing on using models, developing solutions, and engaging in argument from evidence.

A project framed as an engineering design project from the beginning may offer students meaningful opportunities for science learning, especially when they must consider trade-offs in their designs (Purzer et al., 2015). This type of consideration of design trade-offs may be especially useful in helping students to integrate their science ideas with their design decisions. Design projects have been found, in some cases, to positively impact students’ scientific reasoning (Silk et al., 2009). However, these students may not learn complex science concepts if their focus is on incidental aspects of design. Hands-on projects that directly follow a related science unit may allow students more time to focus on understanding the complex scientific phenomena they are being asked to apply, while still motivating them to learn the concepts in order to apply them to their design. However, the separation of the science content from the design project may seem disjointed to students and lead to lower motivation in learning the concepts.

We use knowledge integration assessment items (Linn & Eylon, 2011; Liu et al., 2008) at pretest and posttest targeted at three specific areas to better understand how each of our conditions impacts learning. These items measure science concept integration, engineering design practices, and the integration of science and engineering design practices. While there has been much work done to advance engineering education at the K-12 level (e.g., National Research Council, 2009; Bybee, 2011), there has not been as much work done to develop valid items for assessing engineering practices.

The two conditions in this research are meant to understand two common ways hands-on activities are framed in the classroom. By understanding the benefits of each method of framing, we hope to develop a curriculum that helps students to integrate their ideas about science concepts and engineering design better. While teachers may have their own preferred way to conduct hands-on projects in their classrooms, this work is meant to help strengthen student learning in both the science and engineering domains no matter the framing of the classroom project.

Methods

Participants and procedures
One teacher and her 153 students participated in this study. Out of these students, 139 students completed a pretest, (some part of) the curriculum, and a posttest. The pretest was conducted one day before beginning the unit, and the posttest was conducted one day after finishing the unit. Both the pretest and posttest were administered to students individually. Pairs, or in some cases triads, of students were assigned to collaborative workgroups by their teacher to work on curriculum. Workgroups were randomly assigned to a condition (science or engineering) by the software. All students received the same curricular content, but activity focus and order varied by condition.

Curricular materials
This study was implemented in a curriculum module entitled Solar Ovens in the Web-based Inquiry Science Environment (WISE), which utilizes a variety of instructional and assessment tools (Linn & Eylon, 2011). The goal of the unit was to familiarize students with the way energy transforms from solar radiation to heat through a hands-on project and interactive models, covering the modeling aspect of the Science and Engineering Practices of the NGSS, as well as the standards associated with energy, specifically standards related to the transfer of thermal energy (NGSS Lead States, 2013).

The solar ovens curriculum within WISE has been designed and refined with the collaboration of multiple expert teachers and researchers to help students test and refine their ideas about energy transformation. The curriculum seeks to help students utilize their ideas about how radiation works in various contexts, like in the atmosphere and inside solar ovens.

Students in both conditions followed a modified “design, build, test” approach. An important feature of this unit is a budget activity in which students make decisions about and justify the materials they choose to use for building (Figure 1). During the design phases, students also draw pictures of their ovens and explain how energy transfer will occur. Students also use an interactive model of a solar oven, designed using NetLogo (Wilensky, 1999), to test features in the solar oven and understand how solar radiation transforms into infrared energy (Figure 2). Students generate trials using the model by allowing the model to run for 5 simulated minutes without changing the input variables. When students test their physical prototypes they also test them under a lamp for 5 minutes. After each trial is generated, it is automatically added to a table, allowing students to track
the trials they ran and the results of those trials. The computer model has been previously tested to understand how students use it at different points during the curriculum and how it impacts learning. Our earlier findings indicate that the computer model aids students in integrating their science and design ideas, and that students interacting with the model earlier during the curriculum (during the planning phase) benefit more than students who interact with the model later in the curriculum (the reflecting phase) (McBride, Vitale, Applebaum & Linn, 2016). After designing, students build physical solar ovens, which are tested under lamps with a common set of requirements, so results are comparable between trials and groups.

<table>
<thead>
<tr>
<th>Materials &amp; Costs</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sheet of construction paper (8.5 inch x 11 inch)</td>
<td>$2</td>
</tr>
<tr>
<td>12 inch x 22 inch sheet of tin foil</td>
<td>$7</td>
</tr>
<tr>
<td>12 inch x 22 inch Sheet of plastic wrap</td>
<td>$6</td>
</tr>
<tr>
<td>You can RENT: 12 inch x 12 inch Sheet of Plexiglass (thick plastic)</td>
<td>$10</td>
</tr>
<tr>
<td>8 feet of type (Duct, masking, clear)</td>
<td>$3</td>
</tr>
</tbody>
</table>

Figure 1. Student budget (left) and example of student design (right); students were given $20 for their first design iteration and $13 to add to their oven for the second iteration.

Figure 2. The interactive model used by students for design solar ovens and understanding energy transformation, with an automatically generated table below.

Condition differences

Conditions did not differ in content, only in the order the content was presented and in the framing of questions or activities. In the engineering condition students were introduced to the design project in the first step, then were prompted to learn or consider science concepts during the design process. In the science condition, students learned all the science concepts at the beginning of the project in a module about the atmosphere and were then introduced to the design project as a way to apply what they had just learned. Students in each condition used a concept-mapping tool to map energy flow. In the engineering condition, students mapped energy flow in their solar ovens, while in the science condition students mapped energy flow in the atmosphere. These differences are outlined in tables 1 & 2.

Tables 1 & 2: Table 1 (left) shows the main steps in the curriculum for the engineering condition, including the number of steps. Table 2 (right) shows main steps in the curriculum for the science condition.
The pre- and posttest assessments we used consisted of 9 assessment items. These items fell into three areas: science concepts, engineering practices, and the integration of science and engineering. All items use short response format and are scored using knowledge integration rubrics. Of these 9 items, 5 items measure integration of science concepts, 3 items measure integration of engineering design ideas and practices, and 1 item measures the integration of design practices with science concepts.

One of the science concept items, Car, prompted students to explain what would happen to a car left in the sun during a cold day. In an engineering item, Budget, students were asked to describe how two fictional students would build solar ovens using two different lists of materials and then to describe the tradeoffs made in each design. In the science-engineering integration item, Model, students were asked to use a basic solar oven model (like that shown in Figure 2, but with only a box shape drop-down option) to help a fictional student determine whether a tall, skinny box or a short, wide box would heat up faster. The pretest and posttest were composed of the same items.

While the science and engineering integration items measure how well students link their ideas about design or about science concepts, we were particularly interested in the performance of students in each condition on the integration item, since a goal of this curriculum is to help students use their science ideas to justify their design decisions. This integration item has been tested with over 1000 students in prior work.

We also use the automatically generated table from students’ interactions with the interactive computer model (Figure 2) to analyze how many trials students ran during the design process. In addition, we use three other measures of students’ interactions with the interactive computer model. We use the amount of time students spent on the project step that included the computer model, the number of clicks students made in the computer model, and the average number of clicks made per hour (time spent divided by number of clicks). All of these measures come from analysis of student log files.

Analysis approach
To measure knowledge integration, the items were scored using knowledge integration rubrics to assess links between multiple normative science ideas (Linn & Eylon, 2011; Liu et al, 2008). The knowledge integration rubric for Model, the science/design integration item, shows how links are scored (Table 3). Multiple researchers develop the rubrics for each item; initial scoring of data is also done by at least two researchers, with high inter-rater reliability ($\kappa > 0.8$).

Since this research investigates the differences between framing as a whole project (more similar to engineering) or as an application of concepts (more similar to science), our analysis looks at whether there are
differences between conditions on the science, engineering, or integration items. However, unless otherwise specified, we examine the corpus of all 9 items.

To analyze the differences between conditions based on how students used the interactive computer model during the design phase of the project, we used a count of the number of trials run by each group. Each trial is added to an automatically generated table after students allow the model to run for 5 simulated minutes (takes about 30 seconds to 5 minutes in real time). We do not count trials that were not allowed to run for shorter than this time period because, since they were not added to the automatically generated table, students did not have a record of them and were therefore not able to look back at these trials while making their decisions. This analysis is done at the workgroup level.

Table 3: Sample knowledge integration scoring rubric for the Model pre/post open response item

<table>
<thead>
<tr>
<th>Score</th>
<th>Level</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Off Task</td>
<td><em>I don’t know.</em></td>
</tr>
<tr>
<td>2</td>
<td>Irrelevant/Incorrect</td>
<td><em>David is correct because I chose the skinny and tall one and the heat went up really fast.</em></td>
</tr>
<tr>
<td>3</td>
<td>Partial</td>
<td><em>David's claim is not correct because in the model it show solar radiation stayed trapped inside the wide and short one making heat easily trapped inside.</em></td>
</tr>
<tr>
<td>4</td>
<td>Basic</td>
<td><em>David's claim is incorrect because the skinny box got to 33.8 in 2 minutes and the wide box got to 44.7 in 2 minutes. The wider box could keep a lot of energy because of the space and the skinny box doesn’t have a lot of space. So, this means David was wrong.</em></td>
</tr>
<tr>
<td>5</td>
<td>Complex</td>
<td><em>David's claim is incorrect because the more area for radiation to come the more radiation can get trapped and turn into heat. There is less of the when you have a skinny box.</em></td>
</tr>
</tbody>
</table>

Results

A t-test of pooled pre- and posttest data across conditions revealed a significant effect of testing session \( t(304) = -6.44, p < 0.0001 \), demonstrating that across both conditions students made gains from pre- to posttest (Figure 3).

There were no overall differences between the science and engineering conditions. When considering the groups of science and engineering assessment items, there were non-significant differences in condition differences. Students in the science condition made slightly greater gains on the science assessment items between pretest and posttest and likewise students in the engineering condition made slightly greater gains on the engineering assessment items between pretest and posttest; neither of these differences were significant.

When considering the integration assessment item, there was a significant impact of condition. Using a regression model, students in the engineering condition scored higher on the posttest integration item, when controlling for pretest score \( \beta = 0.18, p < 0.01 \). This is shown in Figure 4.

We also analyze the number of trials students run in the interactive computer model during the design phase of the project. Groups in the engineering condition ran significantly more trials than those in the science condition \( \beta = 0.33, p < 0.02 \). Figure 5 shows data on the variance between the conditions in terms of the number of trials run. In the engineering condition, more of the groups used the model to run trials, and a larger proportion of groups ran more than one trial. In the science condition, many groups did not even allow the model to run for a full trial, and of groups that did run any trials, a majority of them only ran one trial.
There was not a significant difference between conditions when examining the amount of time students spent using the computer model. On average, students spent about 20 minutes using the computer model, with students in the engineering condition spending slightly longer on average than students in the science condition. However, there was a significant difference between conditions when looking at the number of clicks, or actions, students made while using the model ($\beta = 0.34, p < 0.01$), with students in the engineering condition making 25 more clicks than students in the science condition (mean for engineering condition: 56, mean for science condition: 31). Since students in the engineering condition ran more trials, we would generally expect them to also have made more clicks. When combining the measures of time and clicks to be the number of clicks per hour (calculated: clicks divided by time), we find no significant difference between conditions. This measure is important to check because in some cases, students may make rapid clicks on an interactive model without allowing the model to run and reveal the results or patterns to students. We found that there is generally a linear relationship between the amount of time spent and number of clicks in the model. We also found few outliers, meaning that most students were using the model appropriately.

Students used the model to run more trials in the engineering condition, even though students in both conditions generally spent the same amount of time using the model. This may mean that students in the engineering condition used the model more effectively to test their ideas. This is likely because students are introduced to the model very early in the project, so they are using the model to add and test new ideas about their design. Students in the science condition may have already been considering their design throughout the project, but before they were able to test their ideas using the model. This may have caused students to become attached to certain choices they made before they had a chance to use the model to test design options.
Figure 5. left: Histograms showing the number of trials done using the interactive computer model in each condition; right: scatterplot showing the number of clicks and the time spent using the model for each group of students.

Conclusions and implications
Students in the engineering condition were more successful in integrating their science ideas with their oven design than were students in the science condition. Students in the engineering condition may have used the design of their oven as an artifact for testing their science ideas. The students in the engineering condition conducted more trials than students in the science condition giving them more opportunities to test their science ideas. In the science condition the students may have seen designing the solar oven as separate from learning the science concepts.

Students in the science condition spent more of the curriculum solely focused on learning science concepts, therefore it makes sense that students in this condition would do slightly better at integrating their science ideas on the science integration items. Similarly, students in the engineering condition spent a longer time considering the trade-offs of their designs, and also performed better on items that measured engineering practices, like analyzing designs for trade-offs. In addition, students in the engineering condition may have seen the ideas they tested in the interactive computer model as more open to questioning, which may have encouraged students to test more ideas (Sandoval & Morrison, 2003). Students in the science condition seemed to be more attached to their original ideas, testing fewer ideas in the computer model. The engineering condition seemed to open students up to more possibilities in their design, while the science condition in some ways gave students a more limited idea of the possibilities for their designs. Since adding and testing new ideas is a proven feature or curricula that improves student learning (Linn & Eylon, 2011), it is important to emphasize this in all design projects, including those following a format more similar to the science condition.

These results indicate that there are benefits for each type of framing. This is important to recognize in aligning the design of the curriculum with teachers’ learning goals for their students. In this work we recognize that there may be outside factors that impact teachers’ choices in how to frame a hands-on project. However, the results show that there are impacts in what students take away from different framings of the same hands-on project. To improve the science condition, curriculum designers or teachers may have to work to integrate the addition and testing of ideas earlier during the curriculum to overcome students’ fixation on certain ideas during the design process. To improve the engineering condition, science concepts must be emphasized.

This work included only minimal differences between conditions; ordering and question framing on only some curricular activities. While it still generated a useful and statistically significant finding, it would be more helpful for understanding how to frame design projects if the conditions were separated even further. However, separating the conditions further may be very challenging for one teacher to orchestrate (since students are randomly assigned within class periods). In addition, understanding how the framing of hands-on projects impacts learning outcomes also relies on valid and reliable measures for learning. While this has been studied and many psychometrically valid items have been developed in science contexts, this is not yet the case for engineering design in K-12 settings. This work would benefit from further research into measuring engineering and design practices in K-12 settings and the development of useful items that are not reliant on specific scientific content.

References


Negotiating Epistemic Agency and Target Learning Goals: Supporting Coherence from the Students’ Perspective

alizazivic@u.northwestern.edu, jfsmith@u.northwestern.edu, reiser@northwestern.edu, kelsey.edwards@northwestern.edu, mnovak@ccl.northwestern.edu, tara.mcgill@northwestern.edu
Northwestern University

Abstract: A tension in designing classroom learning involves balancing the questions and interests of students with the goals of teachers and standards. One approach to navigating this tension in science classrooms is to simultaneously support and constrain students’ questions about an observable natural phenomenon in the classroom and then take up those questions throughout a unit of instruction. Analyses of students’ questions at the start of a middle school science unit and their responses to surveys throughout the unit suggest that students perceive that their questions are indeed driving learning, suggesting the promise of supporting students’ epistemic agency through co-construction of questions, ideas, and investigations in storylines.

Introduction
A core commitment of education reforms across disciplines is to bring disciplinary practices into classrooms. In science education, the focus is on making learning more meaningful by supporting students as they construct and use knowledge rather than learn about ideas that others have built (Duschl, 2008). An emphasis on scientific practices, like those articulated in the U.S. by the National Research Council (NRC, 2012), shifts the goal of science classroom from learning about ideas to figuring out those same ideas by engaging in practices similar to those used by scientists. This, in turn, shifts the role of students in science classrooms from passive, recipients of knowledge to agentive, constructors of knowledge (Berland et al., 2016). In particular, students engage in the communal construction of scientific understandings through scientific practices such as designing investigations, developing explanatory models, and arguing from evidence (McNeill & Pimentel, 2010; NRC, 2012).
Catalyzing and sustaining a shift from learning about science to doing science requires supporting the development of students’ epistemic agency—students’ involvement in directing and monitoring knowledge building processes (Damša et al., 2010; Stroupe, 2014). In the science classroom, this means that students need to be a part of identifying problems to work on, deciding how to pursue these investigations, and partnering with teachers to reach consensus about what has been figured out (Edelson, 2001; Reiser et al., 2017). This active role facilitates coherence from the students’ perspective, where students are aware of why they are doing what they are doing, as teachers guide them in building ideas based on questions and problems they have identified (Reiser et al., 2017). However, it is one thing for students to be aware of where they are going and why; it is another for them to feel like they have the power to decide where to go next.
Attempts to support students’ development of epistemic agency does not guarantee that students will embrace this potential role. Jaber and Hammer (2016a, 2016b) argue that this essential shift depends on epistemic affect, or the emotions felt as students do science, such as fascination with scientific phenomena, the unease that comes from discovering an inconsistency in a scientific explanation, and the joy in making a successful argument (Jaber & Hammer, 2016a, 2016b). They argue that these emotions motivate students to assume the role of epistemic agents and actively engage in the science classroom. This idea is central to epistemic agency. Creating a classroom where students have epistemic agency depends on students feeling like they can and should be doing the intellectual work of science knowledge construction. Therefore, it is critical not only to look at what students are doing in the classroom, but also to understand how students feel about what they are doing in order to create learning environments where students are truly directing the knowledge building.
At the same time, we are in an era of high accountability. Teachers, curriculum designers, and schools are responsible for addressing particular standards which articulate core science ideas. Thus, there is an inherent tension between supporting epistemic agency for students while also meeting target learning goals. This tension between externally mandated learning targets and student interests is not new, and goes back at least to Dewey (1902). With the articulation of the Next Generation Science Standards (NGSS) in the U.S. (NRC, 2012; NGSS Lead States, 2013), this tension emerges as a fundamental concern.
In this paper, we examine students’ responses to a curriculum unit designed to help teachers negotiate this tension—guiding and supporting students as they meet target learning goals while they simultaneously partner with teachers in managing the focus and trajectory of investigations. The unit uses a storyline approach to optimize these tradeoffs. While each storyline has target NGSS disciplinary core ideas (DCIs) and performance
expectations (PEs) and so the general pathway of the unit is pre-written, each storyline explicitly supports teachers in drawing out students’ prior knowledge, working with students to identify questions and problems, and co-constructing next steps to investigate these questions and problems, engaging students as partners in the knowledge building (Reiser et al., 2017). This study examines students’ experiences in a unit designed to both take up their questions and ideas about what to do and how to do it, as well as to develop core ideas and practices targeted in the standards and curriculum. First, we consider whether the questions students raise in response to observing an anchoring phenomenon align with the unit design. Further, while we, as researchers may recognize that the unit addresses students’ questions, students’ responses to the unit matter as well. Thus, we also examine the degree to which students perceive their epistemic agency, and the consequences of these perceptions for their affective reactions. As students move through the unit, do they see how their questions are being answered? Do they feel they play a role in deciding where to go next? What are their affective responses to these experiences?

**Methods and analysis**

We invited teachers from across the United States to apply to participate in the *Learn While Teaching* project. We selected 27 middle school and high school science teachers from a pool of 86 applicants, 20 of whom agreed to participate in the program. Participating teachers were from urban, suburban, and rural school districts in CT, IL, MA, MI, OK and VT. All of the selected teachers had participated in previous professional development related to NGSS. Before the start of the school year, teachers participated in five days of in-person professional development on supporting NGSS-aligned classroom learning in general and prepared to enact curriculum materials for middle school physical science or high school biology. Teachers were supported during their enactment of the curriculum materials through biweekly virtual meetings.

This paper focuses on data collected from five of the middle school teachers using the Sound storyline unit designed by a team of teachers and researchers. The unit consists of 24 lessons across roughly 7 weeks. Four core questions drive the unit: (1) How is sound created? (2) How does sound travel? (3) How is sound detected? (4) How can technology store and recreate sound? The unit targets NGSS performance expectations that ask students to develop models of waves in terms of amplitude and energy (MS-PS4-1) and that account for the ways in which waves are transmitted, reflected, or absorbed by different materials (MS-PS4-2) and can be transmitted by signals that encode information (MS-PS4-3).

To support coherence from the students’ perspective, the unit organizes learning in a storyline, in which students develop questions from phenomena that drives their knowledge building (Reiser et al., 2017). An anchoring phenomenon helps elicit students’ initial questions, which are then used to uncover related questions from students’ own experiences. At each step, pending questions or gaps students identify become the motivation for the next investigation. In this particular unit, the anchoring phenomenon is a homemade record player that plays music. The questions students generate about the anchoring phenomenon are organized on a driving question board (Figure 1; Singer et al., 2000), a public representation of student ideas (Windschitl et al., 2008).

![Figure 1. Example driving-question board (DQB) from a middle school sound classroom.](image)

We consider two sources of data: (1) photographs of student questions posted on the DQB following the anchoring phenomena, and (2) surveys examining students’ views of their own learning and engagement.

**Student questions**

We analyzed students’ questions from the DQBs from a sample of three teachers’ classrooms to examine how students’ questions aligned with the target science ideas. Analysis of DQBs from additional teachers is ongoing. Examining how the unit takes up students’ questions is an important first step in determining whether the unit supports students in developing epistemic agency. Do students want to go where the unit will be taking them? If students raised a large number of questions that went unaddressed or did not ask questions at all, this could work
against developing feelings of epistemic agency. The DQBs allowed us to examine whether the anchoring phenomenon precipitated questions that could be taken up in the classroom to drive learning.

Each teacher took photographs of their DQBs during the first few weeks of the unit. A researcher transcribed all questions that were visible (i.e., not covered up by another piece of paper) and legible (i.e., focus of photograph was clear as was students’ handwriting). Three researchers, a designer who co-wrote the unit and two researchers who studied the unit in classrooms, coded 265 questions from the DQBs across the three teachers’ classrooms. Specifically, we considered each question in terms of two deductive codes: (1) Question Match, if the question asked was about a specific phenomenon and mechanism addressed explicitly in the lessons of the unit, and (2) General Mechanism, if the question could reasonably be expected to be answered mechanistically by the end of the unit as students have learned the relevant science ideas to be able to do so.

Table 1 includes example questions from two students alongside questions and related phenomena from the unit. Both students’ questions require scientific explanations that address the creation and comparison of different pitches, which are the focus of Lesson 6. The first question, “How can peoples voices be different like deeper and higher for example my pit bull has a deep bark and the Jack Russell has a high voice?” involves a phenomenon, dogs barking at different pitches and ranges of pitch. The specific phenomenon of dogs barking was not addressed in the unit materials, so researchers coded the question “0” for Question Match. However, because the first question about dogs’ barks could be at least partially explained by the general mechanisms uncovered in the unit, we coded the question with a “1” for General Mechanism. Specifically, in Lesson 6 students are expected to develop models for pitch and in Lesson 7 they read about human vocal cords. Therefore, the general mechanisms and even an analogous case (humans) might be enough to support students in reasoning about dogs barking at different pitches. In contrast, we awarded the second question in Table 1 about xylophones and pitches a code of “1” for both Question Match and General Mechanism, because the specific phenomenon of how different length objects, including xylophones in particular, can make different pitches is directly addressed in the unit. Researchers separately coded the General Mechanism and Question Match categories and then compared each of their 265 scores in both categories to arrive at consensus. Discussion often involved returning to the written curriculum materials. Table 2 includes these same questions to illustrate the way the questions were coded. We coded the third question in Table 2, which asks about dogs barking and cats meowing, “0” in both categories.

Table 1: Example questions and corresponding questions and phenomena investigated in the unit

<table>
<thead>
<tr>
<th>Example Student Questions</th>
<th>Related Unit Questions</th>
<th>Related Phenomena Investigated in Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>How can peoples voices be different like deeper and higher for example my pit bull has a deep bark and the Jack Russell has a high voice?</td>
<td>Lesson 6: How do the vibrations from different sound sources compare for higher vs. lower pitch notes?</td>
<td>Guitar string plucked with finger pressing on string at different locations, xylophone bars of different lengths hit with mallet, and music boxes that students wind and play in their hands show patterns in pitch of note vs. length of object that is struck or plucked; patterns in effects on a long, thin wooden stick that we clamp down and strike and then, using a motion detector, measure the position of the end of the stick.</td>
</tr>
<tr>
<td>Why when you hit the larger key on a xylophone the lower the pitch is?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Example questions and corresponding codes

<table>
<thead>
<tr>
<th>Student Question</th>
<th>General Mechanism</th>
<th>Question Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>How can people’s voices be different like deeper and higher for example my pit bull has a deep bark and the Jack Russell has a high voice?</td>
<td>1 (yes)</td>
<td>0 (no)</td>
</tr>
<tr>
<td>Why when you hit the larger key on a xylophone the lower the pitch is?</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>How come a dog barks and a cat meows?</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Student surveys

Understanding the alignment between students’ questions and the structure of the curricular unit is important, but what if students had aligned questions and they cared little about the answers? Or perhaps students might not have believed that their questions mattered and had little confidence they would be answered. Since we argue that epistemic affect is a key part of students developing epistemic agency, we used student surveys to capture students’ thoughts and beliefs about their science classroom. The goal of these surveys was to better understand whether students felt like they had epistemic agency in their science classrooms and to start to unpack what the consequences of those beliefs were. Teachers administered the student survey, designed to take 10 minutes, approximately every other week during the curricular unit. Teachers had access to all of the responses in order to
inform their teaching. Survey questions were adapted from previous work (Penuel et al., 2016). We developed three categories of survey questions that attempted to capture three dimensions of epistemic agency: 1) the intellectual work, 2) the social dimensions of that intellectual work, and 3) students’ affective response to their classroom (see Table 3). While this study is ongoing, this paper examines 1119 completed student survey responses from 373 students from five teachers (21 class periods).

Table 3: Student perceptions survey questions by category

<table>
<thead>
<tr>
<th>Category</th>
<th>Questions</th>
</tr>
</thead>
</table>
| Intellectual work   | [Learning was student driven]                                                                                      On a scale from 1 – 5: “How did you learn today?”  
(1 = “The teacher told us everything we need to know”; 5 = “We figured everything out as a class, with the teacher helping but not telling us the answer.”)                                                                 |
|                     | “I know why we did what we did in class today” (yes, no, unsure)                                                                                      “I figured out something today that helped us make progress on the DQB”  
(yes, no, unsure)                                                                                                                                                                                                 |
|                     | “Do you think your class will figure out an answer to any of your questions”  
(yes, yes – at least part of one, no, other)                                                                                                                                                                                                 |
|                     | “I know where we are going or what we are likely to do next time in class”  
(yes, no, unsure)                                                                                                                                                                                                 |
| Social dimensions   | “Today I shared my thinking out loud: 1) with people in my small group, 2) in a whole class discussion, 3) with people in my group and in a whole class discussion, or 4) with no one” (choose one)                                                                                            |
|                     | “Listening to other students in my group helped me improve my thinking”  
(yes, no, unsure, I didn’t listen to another student today)                                                                                                                                                                                                                  |
|                     | “When other students shared their thinking out loud with the whole class today, I understood their explanations” (yes – most of the time, yes – some of the time, no – didn’t understand, no – didn’t hear, no – no one shared their thinking) |
| Affective response  | “What we did or learned about in class today matters to me, matters to the class, matters to the community, none” (choose any)                                                                                                                                                                                                 |
|                     | “Today’s science lesson made me feel: excited, bored, confused, like a scientist, confident, happy, sad, afraid, angry” (choose any)                                                                                                                                                                                                       |

The first category addresses key aspects of the epistemic work that students were being asked to do in the classroom such as who students thought was driving the learning and whether they believed that they were and would continue figuring out their own questions. The second category looks at the social nature of communal knowledge building: how did students share their own thinking, did they listen to others, and did they find listening productive. The last category focuses on the students’ motivation: did they (or anyone) care about the work being done in their classroom, and what was their affective response to that work.

Findings

Our findings consider the degree to which the unit, as designed, takes up questions students raised at the beginning of the unit as well as students’ reported experiences in classroom.

Student questions

41% of the questions students raised across the three classrooms examined in this part of the study (110 of 265 questions) are directly addressed in the unit design. Specifically, students’ questions may be addressed by an investigation, uncovered in discussions guided by the teacher aimed at figuring out key ideas, and/or addressed in readings in the unit materials. Both the curriculum and teacher provide support for students to connect and reflect on the answers to these questions. (It is an open question about teachers’ actual uses of these materials, a question which we do not take up in this paper but will address in future research.) An additional 21% of students’ questions (55 of 265) can be figured out using a general mechanism that is at the core of the unit. Indeed, there is a range of questions and underlying phenomena that can be explained using the model that students co-create as part of the unit. Overall, 62% of students’ questions (165 out of 265) are directly or indirectly addressed in the unit design.

The findings above support our design conjecture that the opening routine of the unit, which involves students observing and attempting to explain an anchoring phenomenon, connecting to their own experiences, and
uncovering questions they have, can provide support for focusing students’ questions on important learning goals, while also allowing for some freedom and heterogeneity of student questions. Furthermore, questions students raised in response to the anchoring phenomenon span all four sections of the unit: (1) How is sound created? (2) How does sound travel? (3) How is sound detected? (4) How is technology used to store and recreate sound? Thus, students’ questions can be taken up as motivators for each section of the unit.

Some of the unanswered questions required mechanisms not addressed in the unit, such as questions involving the speed of sound (e.g., “What is a sonic boom?” and “Can a plane be faster than sound?”). However, other questions reflect issues for which the unit may help students make partial progress. For example, consider “How do whales use echolocation?” coded as a “0” for both General Mechanism and Question Match. A number of relevant pieces of this puzzle are addressed in the unit—how one type of animal (humans) produces sound, how humans perceive sounds, how sound travels through water, and how sound can be reflected. Thus, students may be able to construct a more thorough (although still incomplete) explanation of this phenomena after the unit. In this analysis, however, we used the more conservative metric of full match on mechanism or specific question.

Some of these unanswered questions may be productive questions to address in high school. However, in some cases, the purpose of NGSS was to cut down on the amount of science topics that would be addressed in school curricula and focus on core explanations that address a broad of phenomena, a concern from even before NGSS (AAAS, 2001). Keeping concerns about unwieldly curricula in mind, looking closely at students’ questions and whether that are answered as part of the classroom work using this unit allows designers to potentially expand and redesign the unit to incorporate questions that seem to be emerging across multiple classrooms in diverse settings. Indeed, we are seeing instances where teachers involved in this study are taking up questions and expanding the unit on their own, presumably to continue to make the unit student-driven.

Our next question is whether students actually recognize that the unit addresses the questions that they have or see how the general mechanism could be used to answer their questions. It is one thing for curriculum designers to assert that students’ questions are addressed in the unit; it is another for students to recognize this or feel that science class matters to them. Students’ responses to surveys provide evidence to address this question.

**Student surveys**

When analyzing the student surveys, we were interested in students’ responses along the three dimensions of epistemic agency outlined earlier and how these dimensions interacted with one another.

**Analysis 1: What were students’ perceptions of their learning experiences?**

When examining the survey questions targeting the intellectual work dimension, we found that many students had similar responses. 93.2% of students (1030 out of 1105) rated their class as a 3 or higher when asked how student directed the learning was in their class (1 = teacher driven and 5 = student driven). 89.7% of students (996 out of 1110) reported knowing why they did what they did in class. 73.6% of students (819 out of 1113) reported that they figured out something in class that would help them make progress on the DQB while 87.9% of students (950 out of 1081) reported that they believed that their class would figure out at least part of their questions. Additionally, 60.1% of students (656 out of 1092) responded that they knew where their class was going.

This pattern continued when examining the questions relating to the social dynamics of the intellectual work. 83.6% of students (929 out of 1111) reported sharing their thinking out loud with either their small group or the whole class, 90.3% (1004 out of 1112) felt that they understood their peers’ ideas at least some of the time, and 79.4% (885 out of 1114) believed that listening to others helped them improve their thinking. Student responses were a little more varied with examining students’ affective response to the unit. 72.0% of students (885 out of 1115) reported a positive affective response to their classroom (excited, happy or confident), 41.1% reported feeling like a scientist, 22.1% reported feeling bored, 16.8% reported feeling confused, and 3.24% reported negative emotions (angry, sad or afraid). Similarly, while 83.5% of students (879 out of 1053) reported that what they did in class mattered to the class, only 61.1% (643 out of 1053) reported that what they did in class mattered to them and only 17.8% (187 out of 1053) reported that what they did in class mattered to the community.

These frequencies of responses, as well as the high agreement, are powerful because they are consistent across almost all of the three categories of questions. Students did not just like the unit, they also recognized that they were being asked to do the intellectual work in the classroom, participated in the class’ discursive knowledge building process, had faith that their questions would be answered by the class, believed that listening to their peers would help them answer those questions, and overall recognized the coherence of the unit. What we can conclude from generally positive self-report responses, however, is limited. A more informative analysis needs to examine whether variation in responses to these questions is diagnostic of meaningful differences in experience. To examine this, we analyze how these responses are related to one another.
Analysis 2: How are students’ perceptions of these three dimensions related?

There are many interesting correlations between different responses, but what we are most interested in is the relationship across categories. Is there a relationship between what students believe learning is like in the classroom (i.e., the intellectual work), how students act in the classroom (i.e., the social dimension), and how students feel about their learning (i.e., the affective response)?

Our analysis revealed a number of interesting correlations between these dimensions. When looking at the relationship between questions about the intellectual work and the social dimension, for example, students who reported that “I know where we are going or what we are likely to do next time in class” were more likely to feel that “listening to other students in my group helped me improve my thinking,” $\chi^2(1) = 20.42, p = 0.000$. Similarly, students who reported that “I figured out something today that helped us make progress on the DQB” were more likely to feel that “listening to other students in my group helped me improve my thinking,” $\chi^2(1) = 45.47, p = 0.000$. These correlations are particularly promising because it shows a direct relationship between who has epistemic agency in the classroom and how students engage in the classroom.

There also is a relationship between the intellectual work and the affective response students have to class. Students were more likely to feel happy, excited, or confident in class if they reported that class was more student driven ($\chi^2(4) = 59.39, p = 0.000$), if they knew where class was going ($\chi^2(1) = 23.78, p = 0.000$), or if they made progress on the DQB ($\chi^2(1) = 31.88, p = 0.000$). These correlations are particularly interesting because students do not always respond positively when they are asked to take on a more active role in the knowledge building of the classroom; in fact, students can become frustrated with their newfound responsibility (Zivic, 2016).

Analysis 3: What questions influenced students reporting that what they did in class mattered to them?

Our ultimate goal in this analysis is to begin to understand the development (or lack of development) of epistemic agency in these classrooms. We selected the response to “What we did or learned about in class today matters to me” as the most relevant indicator that students perceived ownership in their learning. This statement seems most related to the ideas of epistemic agency, in which students see the learning as under their control and serving their and their community’s interests. To investigate the impact of other aspects of the intellectual work, social environment, and affective responses, we conducted a logit regression with the “class matters to me” as the dependent variable, and the other questions as predictors. Note that all independent variables were binary except “Class Learning was Student Driven” which was rated 1-5. Table 4 presents the results of this regression.

Table 4: Effects of intellectual work, social dimension, and affective questions on students’ epistemic agency

<table>
<thead>
<tr>
<th>Outcome: What we did in class matters to me</th>
<th>Coefficient in Log Odds (n = 996)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affective Response</strong></td>
<td></td>
</tr>
<tr>
<td>Feels happy, excited, or confident</td>
<td>1.116*** (0.177)</td>
</tr>
<tr>
<td>Feels like a scientist</td>
<td>0.610*** (0.152)</td>
</tr>
<tr>
<td>Feels confused</td>
<td>0.589** (0.202)</td>
</tr>
<tr>
<td>Feels bored</td>
<td>0.0912 (0.194)</td>
</tr>
<tr>
<td>Feels angry, sad, or afraid</td>
<td>-0.482 (0.443)</td>
</tr>
<tr>
<td><strong>Intellectual Work</strong></td>
<td></td>
</tr>
<tr>
<td>Class learning was student driven</td>
<td>0.283*** (0.0783)</td>
</tr>
<tr>
<td>Know why we did what we did in class</td>
<td>0.0662 (0.235)</td>
</tr>
<tr>
<td>Made progress on DQB</td>
<td>-0.0290 (0.170)</td>
</tr>
<tr>
<td>My class will figure out my questions</td>
<td>-0.256 (0.224)</td>
</tr>
<tr>
<td>Know where we are going in class</td>
<td>-0.161 (0.152)</td>
</tr>
<tr>
<td><strong>Social Dimension</strong></td>
<td></td>
</tr>
<tr>
<td>Shared my thinking in whole class discussion</td>
<td>0.208 (0.154)</td>
</tr>
<tr>
<td>Shared my thinking with no one</td>
<td>-0.290 (0.218)</td>
</tr>
<tr>
<td>Listening to others helped improve my thinking</td>
<td>0.118 (0.194)</td>
</tr>
<tr>
<td>Understood others when they shared</td>
<td>0.124 (0.265)</td>
</tr>
</tbody>
</table>

Note: Standard error in parentheses, * p < 0.05, ** p < 0.01, ***p < 0.001

Interestingly, none of the survey questions in the social dimension had a significant effect on the outcome. While these social dimensions should be present in a student-driven classroom, they are not unique to this environment. That is, working with peers and helping one another might occur in a situation with work that is assigned from external authority rather than a classroom where students feel ownership over their learning.
Among the intellectual work questions, only the student driven nature of the class was a significant predictor of “matters to me” judgments. Surprisingly, knowing why they were doing what they did and making progress on the DQB were not predictive of feelings of ownership. There were however significant relationships between several aspects of students’ affective response and students feeling that class mattered to them. It is interesting that the class mattering to students was not only associated with positive affects (happy, excited, confident) but also with confusion. One possible interpretation is that confusion is more likely with the kind of investment that comes with agency, and perhaps may be seen as a part of learning.

Conclusion and implications

This study has explored an important tension. On the one hand, our education system is now organized around meeting performance standards, which invites the development of common learning sequences. On the other hand, there is interest in encouraging meaningful disciplinary practices that empower learners with epistemic agency, which might suggest allowing learners to choose their directions and focus. This study has explored the promise and challenges of a storyline approach (Reiser et al., 2017), in which curriculum designers aim to support teachers with sequences of phenomena designed to raise questions that help students build the target disciplinary ideas. Initial data from classroom enactments of a middle school unit on sound reveal that the combination of anchoring phenomena, modeling tasks, and productive talk strategies to support rich engagement in science practices (Michaels & O’Connor, 2017) can support some degree of epistemic agency that helps students pursue the target disciplinary learning goals. We examined whether students were able to generate explanatory questions, and whether these questions took them in directions productive for the learning goals. Analyses of students’ questions from the driving question board revealed that students were indeed effective in generating a large number of explanatory questions that spanned the major subsections of the unit, reflecting the major components of the explanatory model targeted in the unit. The majority of students’ questions were within the scope of the explanatory models the unit is designed to help students develop. In the model students develop, students explain how a vibrating object causes particles of a medium to collide and transfer energy to other particles, eventually reaching something that reacts to the vibrating particles (e.g., parts of an ear that detect changes in air pressure or a window that rattles due to vibrations in air). More than half of the students’ questions are indeed explainable by this target model. In separate analyses, we are examining the extent to which students are successful in developing this explanatory model across classrooms and the teaching approaches that support this learning.

The analyses of student surveys then examined the consequences of linking students’ investigations to their questions in the storyline. For example, it was possible that our analyses would uncover the potential connection of the units’ lessons with students’ questions, but that students would not see those connections or perceive them as important. It was also possible that even if students perceived these connections, they would not influence their affective response or feelings of agency. However, we found several strong positive relationships between the degree to which students perceived connections to their questions and their affective responses. The degree to which students perceived their role in the learning (figuring out vs. being told) influenced their likelihood of indicating that “what we did in class matters to me.” Furthermore, these judgments of class mattering to the student were positively associated with positive affective responses (happy, excited, confident), and positively associated with “feeling like a scientist.” There was also a positive association between reporting that class matters to the student and reporting feeling confused. While we do not know the direction of this relationship (do students feel that class matters more to them because they are confused or are they confused because they feel that class matters to them?), there does seem to be a relationship between feeling that class matters to you personally and reporting intense emotions about the learning process. The results also revealed insight into the importance of student perceived coherence (e.g., reporting an understanding of where the investigations should go next)—students who reported knowing “where we are going” were more likely to report positive affects. In general, these findings reveal a relationship between the intellectual aspects of coherence (knowing why we are doing what we are doing), and the affective underpinnings of the work (positive affect, feeling like a scientist).

In summary, we strongly agree with the concern that Sikorski and Hammer (2017) raise that to support science as a meaningful practice for learners, coherence needs to arise from the students’ perspective. We endorse their caution that “premeditated coherence, the kind of coherence that is planned and designed for students, may inhibit students’ learning to seek coherence for themselves” (Sikorski & Hammer, 2017, p. 929). Our goal in developing storylines that are coherent from the students’ perspective is to support teachers in involving students as partners in managing the trajectory of the knowledge building. We agree there is a potential tension for students’ questions to take them in a different direction from the target learning goals or from developing the canonical science ideas targeted in the NGSS disciplinary core ideas. However, we suggest there are strategies to address these challenges. We suggest that professional learning situated in teachers’ enactment of educative instructional materials can provide teachers strategies for developing students’ questions and supporting their students’
engagement in modeling and argumentation. Empirical studies of candidate phenomena can help identify phenomena for use in curriculum materials that can be effective contexts, with appropriate teacher probing, to elicit questions that, if investigated and explained, would help develop the DCIs. The present study suggests that these types of iteratively developed storylines, co-designed with teachers, can help navigate this tension between empowering students with epistemic agency and designing to support common disciplinary learning goals.

References

Acknowledgments
This research was funded by grants to Northwestern University from the Gordon and Betty Moore Foundation, the Carnegie Corporation of New York, and the US Department of Education, Institute of Education Sciences, (Grant Award # R305B140042). The opinions expressed herein are those of the authors and not necessarily those of these foundations or agencies. We are indebted to the teachers and students who participated in this study.
Exploring the Unknown: Supporting Students’ Navigation of Scientific Uncertainty With Coupled Methodologies

Julia Gouvea, Tufts University, Julia.Gouvea@tufts.edu
Aditi Wagh, Tufts University, Aditi.Wagh@tufts.edu

Abstract: Learning to make progress in science entails learning how to make progress in the face of theoretical and methodological uncertainties. In this paper we explore the potential for a coupled system that integrates computational and experimental methodologies to support learning to navigate scientific uncertainties. We present a case study of a group of undergraduate students in a ‘hybrid’ computational-experimental biology laboratory course as they investigated a complex biological system. We describe the nature of their progress: (1) mapping their theoretical understanding of the system and (2) planning and implementing their investigation. We then describe the role of the coupled methodologies in supporting their progress by (1) focusing the goals of their investigations, and (2) allowing for comparison and triangulation. Our results suggest that coupled methodological systems have the potential to provide a balance of resistances and footholds needed to support students in learning how to flexibly engage in scientific practice.

Keywords: uncertainty, scientific practice, computational modeling, epistemic agency

Introduction

“...that is what you are actually supposed to be doing in science...you are supposed to be trying to find things that are unknown, not known.” – Nick, undergraduate student

Learning how to do science means learning how to make progress exploring the unknown. In this exploration scientists encounter uncertainties – about what they know as well as about how to proceed. Scholars in science education have argued that students need practice navigating scientific uncertainties: to learn how to identify and characterize new problems, plan and design experiments, interpret complex datasets, and make and defend theoretical claims (Ford, 2005; Manz, 2015; Metz, 2004; Phillips et al., 2017; Reiser, 2004). When students have opportunities to grapple with uncertainties they can learn scientific practices, and perhaps more importantly, they can learn when, how, and for what purposes to use them. For instance, Ford (2005) compared students who were taught the structure of experimental design with those who had to figure out how to design their own measures. Students who designed measures learned both the basic structure of a valid experimental test (controlling the variables) and how to evaluate experimental designs. When an elementary class studied by Manz (2015) encountered uncertainties in their investigations into a backyard ecosystem, it led to goal-oriented engagement with scientific practices. However, when some of their results were straightforward to interpret, the class engaged more superficially, ignoring potentially important variation in the data. In both of these examples, encountering and responding to uncertainty created a need for students to engage in scientific practices to attempt to understand something they did not already know.

Creating learning environments that allow students to learn how to navigate scientific uncertainties is a design challenge. In particular, it involves balancing opportunities for independent exploration with resources and guidance required to make meaningful progress (Engle & Conant, 2002; Jacobsen & Wilensky, 2006; Manz, 2015; Reiser, 2004). We have taken up this challenge in the context of an undergraduate introductory biology laboratory course. Many undergraduate biology lab courses emphasize learning procedures and demonstrating pre-defined outcomes (Beck et al., 2014). We aimed to create opportunities for students to learn how to conduct scientific inquiries into complex biological systems (Jacobsen & Wilensky, 2006). In considering this design challenge we drew on recent work in science studies that describes the potential for a coupled methodological approach to support progress on the frontiers of science (MacLeod & Nersessian, 2013). In this paper we report on how a coupled system of computational and experimental methods supported a group of students in making progress navigating a complex biological system.

Coupled methodological systems and progress on the frontiers of science

Systems biology is a frontier science in which researchers face multiple forms of uncertainty; the field currently lacks well-established theoretical models, standard experimental protocols, and complete datasets. Nersessian and
colleagues have studied systems biology research laboratories in order to understand how scientists manage complexity and make progress on the frontiers of science. They note that some research labs specialize in either experimental or computational methodologies, while others take an integrative approach.

MacLeod and Nersessian (2013) conducted a case study of a graduate student researcher working in an integrative lab. They describe her use of computational modeling and experimental methods as *coupled* because investigations in each guided and constrained her progress in the other. They claimed that working within this coupled system allowed the student to make progress in two ways. First, having access to two modes of investigation, and therefore two sources of information, allowed this researcher to map her degree of certainty in her emerging understanding of the system. She was able to use output from each method to triangulate against the other, filling in her growing understanding of the system as well as locating gaps and uncertainties. Second, the two systems helped her identify, plan and implement next steps in her investigations. Because her research problem was open-ended there were multiple open questions and multiple possible approaches to addressing those questions. Working both with models and experiments helped her make choices about what to do next, with activity in one realm often suggesting avenues of inquiry to pursue in the other. And importantly, having access to both methods afforded her the flexibility to respond fluidly, changing course when she got stuck or taking the opportunity following unexpected, but potentially fruitful leads.

Our interest in this case was in considering how designing a learning environment that coupled these two methods could help students learn to make similar kinds of scientific progress, expanding their theoretical understanding and planning their investigations, in the face of uncertainties.

**Design rationale: Coupling computational modeling and experimentation**

This study is part of a design-based research project (Cobb et al., 2003) called ‘Hybrid Labs’ to emphasize the integration of computational and experimental approaches.

Drawing on prior research, we conjectured that introducing computational simulations into a biology laboratory course would enhance students’ theoretical exploration of complex systems. By manipulating system parameters students could test and map their understanding of how the system could possibly behave (Svoboda & Passmore, 2013). Dynamic output from simulation runs could draw students’ attention to how system outputs evolve over time as multiple system components interact simultaneously (Hogan & Thomas, 2001). In addition, computational models necessarily entail simplifying assumptions (Wilensky & Reisman, 2006), which we expected could create source of productive tension when compared with experimental systems (Manz, 2015).

We also centralized the role of experimental design, conjecturing that in developing their own method of measuring intended outcomes, students would have to consider the utility and meaning of their proposed measures (Ford, 2005). We further expected that in designing experiments and analyzing data from a complex biological system that students would encounter resistances (push-back from the material system), and that such resistances would create opportunities for students to accommodate their thinking and practices to interpret unexpected results and plan future investigations (Manz, 2015).

Most importantly, we conjectured that it would be critical for students to understand the two methodologies as linked and to use each as a tool to support progress towards their scientific aims. We wanted to avoid unproductive epistemic tensions between the two methods. For example, Smith et al. (1997) described how early in their learning to apply mathematical modeling to biology, some groups of college students constructed models that ignored important aspects of real systems and dismissed such mismatches as unproblematic. Other groups discounted the model, framing it as subordinate to more realistic experimental systems and claiming that the lack of match with reality made modeling useless.

We explicitly designed activities to emphasize interaction between the two methodologies. In this we drew on prior work in K-12 science education by Blikstein and colleagues (2016) who designed activities in which students simultaneously designed experiments and constructed models. They found that these *bifocal modeling* activities allowed students to identify and attempt to reconcile discrepancies between model output and experimental data (Blikstein et al., 2016). We also instructed students to compare output from simulations with experimental data. In addition, we instructed students to use the two methods as tools to plan steps in their inquiry, using the model to direct their experiments and using experimental data to structure how they interrogated the model.

**Research aim**

Our aim in this study was to explore how the coupled system could support progress in the face of scientific uncertainties that are part of investigating complex biological systems. We did so by identifying a focal group of students who made good progress and (1) describing the nature of their scientific progress, and (2) characterizing how the coupled methodological system supported their progress.
Methodologic approach

We studied a group of four first-year undergraduate students, Nick, quoted above, and his group mates Abram, Damian and Walt, chosen because initial observations showed them using the computer simulation and experiment in coordination. For example, the group spontaneously turned to the simulation to overcome an obstacle in their experimental design. Our analysis of this group as a case study contributes an understanding of how a coupled methodological system can possibly support scientific progress.

Study context: Investigating variable mutation rates in bacteria

We studied the activity of the focal group during the first unit in a three-unit sequence of hybrid labs. In this unit students investigated the question, *under what conditions is it advantageous to have a higher or lower mutation rate?* This question was chosen because the relative advantage to organisms with different mutation rates is a complex outcome determined by organism-level parameters, population-level dynamics and environmental conditions, ensuring that the space of possible outcomes would be large enough that students would not be able to predict outcomes with certainty.

Students could investigate this question using a NetLogo computational model in which they could manipulate the relative mutation rates of two simulated bacterial strains, probabilities of different mutation effects (e.g. lethal vs. beneficial), and environmental conditions (e.g. presence/absence of antibiotic) and observe effects at the population level. Students could also design and conduct experiments with real bacteria by growing high and low mutating strains of *E. coli* in various growth media.

To position the two approaches as coupled we structured the lab activities so that students moved back and forth between the two methods multiple times: first conducting a whole class experiment, then exploring the simulation, then designing their own experiments and analyzing their results, and finally returning to the simulation again. Transitions between the two methods were explicitly framed to support links. For example, the instructor, a member of our design team, encouraged students to use the computer simulation to inform their experimental design and specifically highlighted how they could “manipulate things in the simulation that you can’t manipulate in the experiment.” During analysis of experimental results, he instructed the class to “try to model something that is relatively close to the experiment that you did.” And again the instructor focused their attention on how the simulation could allow them to “watch patterns for longer” and “relate that back to their experimental results.” Finally, for their final presentations and reports, students were instructed to include representations of both simulation output and experimental data and discuss consistencies or inconsistencies between the two.

Data collection and reduction

We collected video of the focal group working on all lab activities, including screen capture of their activity in the computer simulation. We also collected copies of students’ three writing assignments for this lab unit, assigned weekly. Finally, we conducted interviews with three of the four focal students after the three-week unit had ended. In these interviews we asked students to reflect on their experience in the lab, specifically asking them to describe moments of challenge, surprise and engagement.

For this analysis we selected two central episodes because they featured interactions between the computational model and experiment. *Episode one* begins at the start of the second week of the unit and covers the time the group spent designing their experiment, including their spontaneous revisiting of the simulation (2 hours of activity). *Episode two* takes place in week 3 and includes the group’s analysis and presentation of simulation output and experimental results (1 hour of activity). We focused on these in-class episodes, written work and interviews in our analysis.

Analysis

In the first phase of analysis we characterized the nature of the progress the group made in investigating the biological system. Drawing from descriptions of the progress in MacLeod and Nersessian’s (2013) study and emergent patterns in the data, we defined two broad dimensions of progress. The first concerned how students mapped their theoretical understanding of system behavior. Progress in this dimension included: identifying parameters or conditions that could affect outcomes, making specific claims about the possible causes of system behavior, and characterizing the relative certainty of such claims. The second concerned the group’s progress in their investigative actions. Progress in this dimension entailed making decisions about directions of inquiry to pursue as well as articulating the purpose and expected epistemic value of such decisions.

In the second phase of analysis we focused on understanding the role of the coupled methodological system in this progress. Within the two episodes, we identified moments when students were discussing or working with both methodologies. We then characterized how students were using the two (e.g. to compare...
patterns, identify discrepancies, frame decisions in the other) as well as how they talked about the relative certainty of the knowledge they gained from each.

Findings

Progress during episode one: Articulating a question

In this section we describe how the focal group makes progress mapping their understanding of the system as well as in planning their investigation as they work to articulate a question to investigate in their experiment.

Mapping the theoretical space

Prior to any investigation into the simulation or experimental design, the students participated in a whole class discussion that elicited initial ideas about the conditions under which high and low mutators would be expected to have an advantage. For homework they individually wrote a paragraph on these initial ideas. An idea that arose from both was the expectation that high mutators would have an advantage in unstable, novel or stressful environments, whereas low mutators would have the advantage in stable environments. This idea was present in all four of the initial assignments for this focal group of students.

By the end of this episode the theoretical space that the group was considering had expanded (Table 1). First, they had differentiated between two different types of “novel” environment: one in which the effects on bacteria are negative (antibiotic) and one in which the effects are potentially positive (lactose, a new nutrient). They described their understanding of the negative antibiotic environment as more certain: According to Damian, in antibiotic environments, “you mutate or you die.” However, they were less certain about the role of mutation in an environment with a novel secondary nutrient. In interaction with the instructor, Damian reported that he was, “not sure if will impact it that much, because there's not like negative force acting to reduce the population.”

Part of their progress involved characterizing this uncertainty. One possibility, offered by Damian was that the higher mutator might have an advantage if it could “use the lactose that nobody else is using.” But Nick pointed out that they, “don’t know how much they’re going to compete for glucose. ‘Cause like it seems like there should be enough glucose to go around.” By the end, Nick summarized that they should expect the high mutator to have an advantage only if the population is at “carrying capacity” where a large population size would mean competition for resources. In unpacking their uncertainty they articulated the role of resource dynamics and competition in considerations of when a mutation can be considered beneficial (Table 1).

In addition, they began to consider how underlying parameters – the probability that a mutation has a harmful versus beneficial effect on the organism – could influence the relative costs and benefits to high and low mutators. By the end of the episode they put a tentative stake in the ground: if the probability that a mutation results in a metabolic benefit is high, then the higher mutator should have the advantage. However, they also expressed uncertainty about this possibility. For example, Damian wrote in his second assignment,

the rate of deleterious mutations compared to the beneficial metabolic mutations in the E. coli is unknown. Therefore, it is possible that the rate of deleterious mutations could be higher which would result in more [of the high mutating] E. coli dying than gaining the potential benefits from the Lactose metabolizing mutation.

Planning their investigation

Students had the choice of designing an experiment that involved growing bacteria in media containing either an antibiotic or a novel nutrient, lactose. At first the group leaned toward an experiment involving lactose, in part because they wanted to do something different from the whole class experiment they just completed (which involved an antibiotic). Early on, Damian expressed concerns that their experiment was “vague” and lacking a “good theory,” and Nick threw his paper and put his head on the desk. Their discomfort seemed related to uncertainty about the outcome of their experiment described above: They were unsure if the high mutator would actually benefit from being grown in lactose.

By the end of this episode the group made progress in planning their investigation and convincing themselves of its value (Table 1). This is evident from their shifting descriptions of their research question. Early iterations of the question were phrased in terms of desired output. They asked whether or not the higher mutator would have an advantage and seemed frustrated when they were unsure about this outcome. Later the group shifted to problematizing the role of lactose in the system. First, they asked whether or not the ability to digest lactose should be considered a benefit. Second, if digesting lactose were a benefit, would it be enough of a benefit to allow the high mutator to outcompete the low mutator? This more elaborated research question seemed to shift the group from a focus on outcomes to understanding what their experiment might tell them about the system.
writing down this final version of the question, Damian explicitly expressed his confidence in the research question: “I understand what we are doing. It's just…is it going to create a difference? But that's like part of the experiment I guess.”

Table 1. Summary of Progress in Episode One

<table>
<thead>
<tr>
<th>Progress</th>
<th>Beginning of Episode</th>
<th>End of Episode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping the Theoretical Space</td>
<td>• Predictions based on broad categories (unstable vs. stable)</td>
<td>• Differentiate between novel environments with positive (new resource) and negative (antibiotic) types of selective pressures</td>
</tr>
<tr>
<td></td>
<td>• Adaptation by high mutators as the primary mechanism of advantage</td>
<td>• Identify importance of and uncertainty around resource dynamics and competition</td>
</tr>
<tr>
<td></td>
<td>• Differentiate between novel environments with positive (new resource) and negative (antibiotic) types of selective pressures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Identify importance of and uncertainty around resource dynamics and competition</td>
<td>• Begin to attend to role of parameter values</td>
</tr>
<tr>
<td>Investigative Actions</td>
<td>• Initial question is too “vague”</td>
<td>• Research question problematizes role of lactose in benefiting the higher mutator</td>
</tr>
<tr>
<td></td>
<td>• Research question as confirming advantage of high mutator</td>
<td>• Confident that question is “good”</td>
</tr>
</tbody>
</table>

Role of the coupled methodological system in episode one

During this first episode, mapping between the entities and relationships represented in each system helped the group locate and characterize the problem they wanted to investigate. The computational model did not explicitly represent nutrient resources, while in the experimental system the group had to choose a nutrient medium in which to grow the bacteria. This mismatch was due to a simplifying assumption made in the computational environment that “metabolic benefit” mutations gave all bacteria an energetic benefit regardless of resource levels (which were not represented in the simulation). In the experimental system, this benefit would depend on the relative abundance of real nutrients.

Prior to experimental design the group had set the probability of metabolically beneficial mutations to high in the simulation and observed that the high mutator had an advantage. This outcome matched their expectation that high mutators should have an advantage in a “novel” environment. Their uncertainty about the role of lactose in the experimental system caused a temporary disruption during which Damian raised concerns about the experiment being “too vague.” In the midst of this frustration Damian suggested that they revisit the simulation. Once again the simulation with metabolic benefit set to high showed them the high mutator winning out, but this time the group questioned that outcome. Watching the graph, Nick wondered, “But is being able to digest lactose a metabolic benefit?” and Damian responded, “That is our question!”

Now the group was using the simulation to mark one potential outcome of the experiment – what they would expect to see if they assumed lactose did provide a benefit to the high mutator. Whether or not they would actually see this result would depend upon whether or not the experiment was “like their simulation” and specifically whether or not the assumption that digesting lactose was beneficial held in the experimental setup. This comparison helped them see their question as both viable and interesting to explore.

Overall, the role of the simulation during this first episode was to represent what the group expected to see. In isolation it functioned to simply to confirm their intuitions; but in interaction with the experimental system, the simulation represented one possible outcome based on a particular set of assumptions. The role of the experiment was both as a source of resistance, causing the group to notice and consider the assumptions in the simulation, and as a promising method to make progress – the results of the experiment could potentially tell them something about the lactose problem they had identified.

Progress during episode two: Analysis and interpretation of results

In this episode the group makes tentative theoretical claims and identifies additional uncertainties as they interpret their results and consider possible next steps in their investigation.

Expanding the theoretical space

At the end of the first episode the group was wondering about resource dynamics. They had also begun to discuss the role of harmful mutations, but had not systematically discussed their impact on the outcome. During this episode they further expanded their map of the theoretical space (Table 2).

One way they did this was by articulating their lack of knowledge of the probabilities of different mutation effects. This is reflected in their final assignments. For example, Nick wrote that while they “could set the mutation type frequency sliders to whatever we wanted” in the simulation, “we had no way of knowing the actual mutation frequency for each different type of mutation in bacteria.” Similarly, Damian qualified claims
about relative success in his report, adding that it would “depend on the ratio of the rate of lethal mutations to the rate of beneficial Lactose metabolizing mutations.”

While they did not know the exact values, by the end of the episode, all members of the group made tentative claims about the relative impact of the parameters. In his final report, Walt argued that, “The lethal mutation rate has a much stronger impact and is a stronger indicator of the high mutator’s likelihood to survive than the beneficial mutation rate.” Abram wrote that, “a higher metabolic-benefit rate would not counteract the effects of the lethal component on the high mutator.” Nick even used this argument about relative impacts to claim that it was unlikely that beneficial mutations are more frequent in the real system because otherwise they would have seen a signal of this in their experiment.

Finally, there was evidence, both during the group’s interaction in the simulation and in their final reports, that they were beginning to attend to mutation dynamics – the rate of spread of mutations through the population. As Abram and Walt ran trials in the simulation, Walt noticed an unexpected result: the high mutator was doing well at first, but after some time the low mutator began to take over. Walt pointed out that this was happening as the percent of beneficial mutations “plateaued” for the higher mutator. In his final report, Abram described how despite the high mutator’s early advantage, “eventually, the low mutator strain gained a beneficial mutation that allowed it survive while the high mutator strain died out.” These observations demonstrate a beginning understanding that advantage can shift over time as different mutations sweep through populations at different rates.

**Planning simulation trials and interpreting and revising experiments**

During this episode, Abram and Walt spent 40 minutes in the simulation and conducted six trials (Figure 1). With the exception of one trial suggested by the TA, Abram and Walt negotiated and defined a rationale for each, using the experimental setup to guide them. Their focus was on comparing the relative impact of lethal and metabolic benefit mutations, which they did systematically. By trial three Walt proposed that, “it seems the low mutator does best when the lethal rate is equal to or higher than metabolic benefit.” In the remaining trials they edged back towards the parameter values in the first trial, increasing the metabolic benefit mutations relative to lethal to find a combination that would allow the high mutator to have the advantage. At the end they summarized across the set, Walt claiming, “I guess that makes sense. If it’s a high mutator, it needs a really low chance of a lethal mutation in order for it to really thrive,” and Abram nodding in agreement.

![Figure 1](path_to_image.png)

A second type of progress was in interpreting their experimental results, which was difficult due to low numbers and an inability to distinguish between strains on some of the plates. Still, when asked by another student during their final presentation if they still believed their initial predictions, they were able to make some tentative claims. Damian responded, “If you look at the simulation, it shows that we saw that the metabolic mutations didn't have a huge impact on the colonies. So I think if we did the experiment again we wouldn't get numbers that showed that the lactose actually helped the high mutating strain.”

Finally, there was progress in identifying new avenues for research. In interviews at the end of the unit, Abram, Damian and Nick all expressed a continued curiosity about the system. Unprompted, Nick articulated two potential ideas for future experiments. He described wanting to “try an experiment in which the glucose was scarce, as that may be a better way to create a selective pressure for lactose digestion where we would more likely see [the higher mutator] benefiting.” Then, in his final report, he proposed an experiment in which he would...
incubate the bacteria for longer, noting that in the simulation it took hundreds of generations for the high mutator to benefit, whereas they only let their experiment run for 30 generations.

Table 2: Summary of Progress in Second Episode

<table>
<thead>
<tr>
<th>Progress</th>
<th>Beginning of Episode</th>
<th>End of Episode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping the Theoretical Space</td>
<td>• Begin to attend to role of parameter values</td>
<td>• Problematize parameter values in experiment</td>
</tr>
<tr>
<td></td>
<td>• Uncertainty around resource dynamics and competition</td>
<td>• Make claims about relative effects of different parameter values (lethal stronger than beneficial)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Attend to mutation dynamics</td>
</tr>
<tr>
<td>Investigative Actions</td>
<td>• Use simulation to confirm one expected result</td>
<td>• Use simulation to systematically explore parameter space</td>
</tr>
<tr>
<td></td>
<td>• Uncertainty over how to interpret results</td>
<td>• Make prediction about what results would have been (without experimental error)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Propose future experiments</td>
</tr>
</tbody>
</table>

Role of coupled methodological system in episode two

During this second episode we identified three ways in which the simulation and experiment interacted to support progress. First, Abram and Walt used the experiment to frame their investigation in the simulation. As Abram described in his final report, in the simulations we ran, we made changes to two conditions that we thought would tell us more about the results of our experiment: metabolic-benefit rate and lethal rate. Since we were not entirely sure whether the high metabolic-benefit rate actually aided the high mutator in our actual experiment (due to not being able to differentiate between the two strains), we decided to run the experiment again.

Abram described thinking that the simulation could give them information relevant to their experiment (to essentially allow them to “run it again”). Further, it focused their choice of parameters to manipulate, ultimately allowing them to claim that the impact of lethal mutations was stronger than the impact beneficial mutations.

Second, the ability to explore parameter space in the simulation caused some group members to see the whole system as more complex. Abram had initially felt relatively certain that the high mutator would win in the lactose experiment. In an interview at the end of the unit, Abram described how, “the fact that we got so many outcomes from the one that we predicted kind of just contradicts our original predictions and kind of shows that …we were really fishing for a specific thing, when it was likely that a lot of possible outcomes could have happened instead.” This is different from how the simulation functioned in episode one to represent simple outcomes. Now both the simulation and the experiment emphasized that the outcomes of the system can vary depending on the conditions and parameter values.

Third, there is evidence that mapping between simulation and experiment influenced at least one of the future experiments proposed by Nick. In his interview Nick described noticing that the simulation ran through hundreds of generations compared with 30 generations in the experiment. Nick noticed that in the simulation sometimes the dynamics would “switch” if run for long enough. These considerations seem to have inspired Nick’s idea to run the experiment for longer, which he explicitly connects to his wondering about whether more time was needed for the beneficial mutation to sweep through the high mutating population. In this example, the simulation once again functioned to open up the space of possible ways the experimental system might behave, this time drawing attention to the importance of mutation dynamics.

Discussion

Designers of authentic scientific learning environments have recognized the need to allow students to explore unknown terrain while at the same time pointing to footholds that will help them make progress. This design tension is often discussed in terms of balancing opportunities for students to encounter problems and resistances on the one hand and access to resources (guides, information, instructor support) on the other (e.g., Engle & Conant, 2002; Reiser, 2004). In this case, each approach – computational simulation and experiment – functioned at times as a source of knowledge and at other times as a source of uncertainty. While the computer simulation was initially used to represent straightforward theoretical predictions, it later became a method for exploring how dynamics changed over parameter space. In the experimental system, the group was initially uncertain about the resource dynamics, but became more certain, despite small numbers, that the experimental system would not favor
the high mutator as they had initially expected. This suggests that an alternative to providing students with targeted help is to provide them with alternative avenues for continuing their investigations.

The progress made by this group was clearly linked to their taking up the idea that the two methods can and should interact; but not all groups understood the system as coupled. This raises a new tension in our design: How much to structure the two methods as interacting as opposed to providing students with access to both methods to use flexibly. In this case, moves between experiment and simulation were primarily guided by the activity structure and framed by the instructor. One exception was the group’s spontaneous move to revisit the simulation during their experimental design. In this moment they demonstrated the kind of flexible use of scientific tools seen in the systems biology PhD student who described the autonomy and confidence she developed in her work: “I like the idea that I’m building my model things are popping up in my head oh wow this would be a good experiment. I plan out the experiment myself and then go into the lab and I do it.” (MacLeod & Nersessian, 2013). Ultimately, the ability to plan and make goal-directed moves in the face of uncertainty is the kind of expertise we want to develop in science students, and coupled methodologies may be a fruitful way to structure opportunities for students to practice this.

References

Acknowledgments
We gratefully acknowledge financial support from the Davis Educational Foundation established by Stanton and Elizabeth Davis after Mr. Davis’s retirement as chairman of Shaw’s Supermarkets, Inc. to J. Gouvea and the Gordon and Betty Moore Foundation through Grant No. GBMF3475 to D. Hammer, Tufts University.
Science Engagement and Identities in Everyday Family Life

Dana Vedder-Weiss, Ben Gurion University of the Negev, vedderwe@bgu.ac.il

Abstract: This self-ethnography investigates what science engagement looks like in everyday life of a "science family" and how science identity emerges through such engagement. I systematically analyze recordings made over a year of science engagement in one family, showing how science was infused in many aspects of its life. However, whereas this engagement supported the development of a science-person identity for one child, it worked to develop a science-antagonist identity for the other. To explore how positioning and roles may help elucidate such local variation, I zoom in on the moment-by-moment interaction in one illuminating event. The analysis reveals how repeating identification within everyday family interactions can help explain differences in identity development. It suggests considering informal science learning environments also as leading to alienation from science and exploring how equal access to science is denied in subtle ways that go beyond socio-historical categories.

Keywords: identity, roles and positioning, family learning, informal science learning, parent-child interaction, self-ethnography, socio-cultural theory

Introduction

Scholars have been increasingly studying the development of a science-related identity with an eye to how young people develop aspirations for a science career, and how they persist in or drop out of the so-called science pipeline (e.g., Calabrese Barton et al., 2013; Carlone, Scott, & Lowder, 2014; Hazari, Sonnert, Sadler, & Shanahan, 2010). Research has looked into ways such identity is shaped in school (e.g. Brickhouse, Lowery, & Schultz, 2000), informal settings (e.g. Calabrese Barton & Tan, 2010) and across contexts (Bricker & Bell., 2014; Zimmerman, 2012). That research predominantly attributes variation in science identity development to socio-historical constructs, such as race, gender, and ethnicity and it often demonstrates how families and other informal learning environments afford the construction of science identity. Rarely does research look into the ways informal settings can also lead to alienation from science. Taking a self-ethnography approach, this study focuses on one family, aiming to: (1) provide a rich account of everyday science engagement in a science family (i.e., a family rich in science habitus); and (2) shed light on how science identities (a science person and a science antagonist) emerge through science engagement in such a family.

Theoretical framework

Identity

To study science identity in family life, I draw upon approaches to identity as fluid, situated, and constructed through activity and especially talk (e.g. Holland, Lachicotte, Skinner, & Cain, 1998; Wortham, 2006). These approaches broadly define identity as "the social positioning of self and other" (Bucholtz & Hall, 2005, p. 586), maintaining that what people do, through language and other activities, is a way of constructing who they are. Identities are constructed in a process of situated learning that involves participation in practices of knowing, talking, doing and being and thus becoming part of a community of practice (Lave & Wenger, 1991). The way novices participate in the community's practice, accept, reject or ignore its practices, and the way others respond to that, shape their identities. Thus, people are identifying and being identified as members or non-members of a community (or: as a "kind of person", Gee, 2000, p. 99), through their interaction with others.

Identity is thus a social and cultural construct emerging as a product of interaction, and not an internal psychological construct that is the source of interaction. Identity is co-constructed by a person and the people with whom s/he interacts (Gee, 2000). This perspective highlights the temporary interactional positions and roles participants play (e.g., the collaborator or the antagonist) or are assumed to be playing. Self-identifying includes playing particular roles of who one is and who s/he wishes to be. Other-identification includes recognizing these roles or assigning other roles. Whereas identity "can change from moment to moment in the interaction, can change from context to context, and of course, can be ambiguous or unstable" (Gee, 2000, p. 99), repeated positioning and roles accumulate through time to stabilize a person's identity vis-à-vis a certain community, its members and practices (Calabrese Barton et al., 2013).

Science identity
I conceptualize a *science person* identity as "positioning oneself (deliberately or not) and/or getting positioned as a “good” science participant" (Carlone et al., 2014, p. 839). I view a science identity not as what someone says about her attitudes to, abilities in, or aspirations in science, nor as simply her feelings towards or actions with regards to science. A science identity emerges in a local setting, constrained or afforded by available resources, constructed by an individual's performance and the way s/he is recognized as a science person by meaningful others (Carlone & Johnson, 2007). Science identity is both situationally emergent and cumulative across time and context. Through years of engagement with science, people develop patterns of positioning and participation. Calabrese Barton and colleagues (2013) named this development "identity trajectory" (p. 65). A growing body of research indicates that science learning trajectories are determined by the age of 14 (e.g., Tai, Liu, Maltese, & Fan, 2006) circumscribed by socio-historical constructs, such as, race, ethnicity and gender (Archer et al., 2010; Brickhouse & Potter, 2001; Carlone, 2004).

Research predominantly demonstrates how school science constrains identification with science whereas informal settings afford it. For example, Carlone et al (2014) showed how school disrupted the science identities of three adolescents whom she followed from 4th to 6th grade. In contrast, Calabrese Barton and Tan (2010) illustrated how a voluntary science program afforded opportunities for urban youth to position themselves as community science experts. In a later study, they (Tan, Calabrese Barton, Kang, & O'Neill, 2013) showed how out-of-school settings provided STEM minded girls with a wide variety of resources and positioning, affording them with opportunities to identify with science, in ways that were not available in the school context.

**Science identities in family life**

Family plays an important role in shaping students’ engagement in, aspirations towards, and identification with science (e.g., Aschbacher, Li, & Roth, 2010; González, Moll, & Amanti, 2006; Bricker & Bell, 2014; Zimmermann, 2012). Archer et al (2012) used the concept of "family habitus" (drawing on Bourdieu, 1984, 2001 and Bourdieu & Passeron, 1979), that is: the family's values, resources, everyday practices and identifications, to explore the extent to which families construct a collective relationship with science and the extent to which this is shaped by their possession of particular sorts of economic, social, and cultural capital. They examined how the everyday family "landscape" shapes, constrains, or facilitates engagement in science, making science aspirations more thinkable for some children than others. Based on interviews and surveys, they found that while a family’s social structural location (e.g., their ethnicity) was important, family attitudes to science and their encouragement and fostering of science in their everyday life had a greater influence on student science aspirations. In the families that they characterized as "science families", science was prevalent in everyday life; children were provided with opportunities, resources, and support to develop a practical ‘feel’ and sense of mastery of science as well as a perception of science as desirable. Many of the parents themselves held science degrees and/or were working within science-related fields. Many of these families held a sense of science being "what we do" and "who we are". Rather than "just another subject", science diffused into all aspects of family life, including daily conversation, and family leisure activities. Nevertheless, in some families, "despite a strong ‘push’ toward science from their parent/s, children did not seem to express a correspondingly high personal identification with (or aspirations toward) science" (Archer et al., 2012, p. 15). This study did not, however, look into the ways in which such "incompatible" identities develop. For example, do children in such families fail to develop "correspondingly high personal identification" and/or do they develop antagonism towards science? Does this happen "despite a strong push" or perhaps because of this push?

While scholars recognize the important role families play in children science identity development, we lack accounts of how everyday family science engagement affords and constrains science identity development (for exceptions see Bricker & Bell, 2014; Zimmerman, 2012, Calabrese Barton et al., 2013). The current study contributes to this line of research by exploring how science identities emerge in a science family through everyday engagement with science. This exploration is based on intensive yearlong participant-observation data collected through self-ethnography.

**Research objectives**

The objectives of this study are twofold:

1. To provide an account of what science engagement may look like in everyday life of a science family.
2. To explore how science identities are shaped through science engagement in such a science family.

**Methods**
Direct observations of family everyday interaction around science are not easy to obtain, particularly interaction at home, as it spontaneously instigated, unfolds and dissolves. To meet this challenge, in this study I took a self-ethnography approach. I observed and recorded family science interaction throughout one entire year in my own family. Self-ethnography is a study in which the researcher describes a cultural setting which s/he is a “natural” part of. S/he works or lives in the setting, using the experiences in, knowledge of and access to empirical accounts to study the setting (Alvesson, 2003). Previous self-ethnographies by a parent-researcher demonstrate the benefits of such an approach as well as some of its challenges (Vedder-Weiss, 2017; Long, 2004; Yoon, 2012). Whereas in ethnography the researcher usually needs to “break in” to the lived experiences of the “natives” and “make the strange familiar”, in self-ethnography s/he needs to “break out” of the familiar, implicit, and taken for granted and make the familiar strange (Alvesson, 2003). To develop such analytic distance, I used linguistic ethnographic methods (Rampton, Maybin, & Roberts, 2015). In addition, I shared the data and my analysis with my family (providing a member check) as well as with colleagues who acted as critical friends.

I acknowledge that the involvement of this study in my family life may in itself have invoked certain behaviors, affecting the interaction and the children's identity formation. However, I suggest that the emphasis this study might have induced on science in my family is aligned with the study objectives: to study science engagement and identity formation in families with a "strong push towards science" (Archer et al., 2012).

Context and data collection
My family consists of a mother (me) who is a science educator and an educational researcher (43 years old at the time of data collection), a father (55 years old) who is a plant biologist, and three boys aged 15, 11 and 8.5. The data for this study was collected during the year 2012-2013 when we lived in Australia for one year due to a sabbatical leave. We usually live in Israel. Throughout this year, I audio recorded events where family members engaged with science content or practice. I regularly had available an audio recording device, which I could switch on when noticing science engagement. I acted as either an-participant-observer or as a mere observer (e.g., when the children were playing among themselves and I have placed the recorder close to the m without physically attending). In total, I audio-recorded 305 events amounting to a total of 26 hours and 52 minutes (with events ranging from 30 seconds to 66 minutes each). In addition, I took field notes, commenting on the audio-recorded events or describing events which I did not audio record. Family members were (and are) aware of the study and repeatedly expressed their active consent to participate.

Analysis
Starting with an exploratory phase, I reviewed the entire data corpus, listening to the audio-recordings, going through my field notes, and writing for each event a research memo summarizing the flow of affairs, including participants, the setting, scientific content or object, scientific and engineering practices (drawing on the NRC 2012 framework for K-12 science education), materials and tools, disciplinary affect (drawing on Jaber and Hammer, 2016), artefacts and any other features that attracted my attention.

I then reviewed the research memos, searching for events that appear illuminating in terms of differing patterns of participation between the two children or in terms of positioning, recognition, and roles. I repeatedly listened to these events, using linguistic ethnographic methods, trying to make sense of what was happening and what may explain differences in participation (Rampton, 2007). Then, I used micro-analytic methods to analyze the sequential unfolding of episodes, which included proceeding slowly through the transcript, asking at each line “What is the speaker doing?”; “Why now?”; “How does this turn of talk respond to what proceeded it?”; “What else might have been done here but wasn’t?” and so forth. I examined the ways in which the children participated and the ways in which others recognized and positioned them. I paid attention to roles, conflicts, power relations, and bids for recognition and floor. At various stages of the analysis, I consulted with family members and colleagues and incorporated their perspectives on the data as well. The original language of the episode is Hebrew, and I have worked from the Hebrew recording and transcript throughout the analysis.

Findings
The analysis revealed that over the year the family engaged with science in a wide array of settings, from designed settings, such as science museums, to un-designed settings, such as playing in the backyard (Vedder-Weiss, 2017). Scientific content varied significantly, including all STEM domains, focusing on abstract objects (e.g., evolution) as well as on concrete ones (e.g., a Koala). While differentially distributed, family members expressed a wide array of disciplinary emotions, including pleasure of having ideas, surprise, pride, confusion and frustration. They engaged in various scientific and engineering practices, such as modeling (e.g., when designing rubber guns), asking questions (such as why didn’t Jack Sparrow’s head get wet when he went into the water with a canoe on his head [in the movie Pirates of the Caribbean]), and planning and carrying out investigations (e.g., investigating...
seedlings growth rate). Such investigations were often accompanied by argumentation, reasoning, and information evaluation. Interestingly, the family usually used mundane tools and materials, such as rocks they collected outside or rubber they bought in the store. Almost never did they use laboratory tools and materials or science kits, although these were accessible to them. The family produced apparatus such as a water trap, image artifacts such as videos, and texts such as letters.

Throughout the varied science engagement described above, Yoav (8.5 years old) often exhibited engagement behaviorally, cognitively, and emotionally. He usually appeared more interested and engaged in science conversations and experimentation than Shahar. Shahar (11 years old), on the other hand, appeared to quickly lose interest, and often attempted to terminate scientific engagement. For example, on April 16 the family joined Dad on a visit to his workplace. Walking around the research institute, they discussed the fall of autumn leaves and how the connection of the leaf to the stem weakens as it changes its color. Yoav said "Mom, look, I picked a yellow leaf. It detached real easily. A green leaf - much harder". Mom confirmed Yoav's observation. Shahar joined them only after Mom encouraged him to do so. Later, they joined a pea's DNA extraction activity offered by a small discovery center. After the activity, Dad suggested to further concentrate the extracted DNA using his lab centrifuge. Mom, Shahar, and Yoav waited outside while Dad went into the lab to centrifuge the DNA. Yoav looked through the window trying to see what Dad was doing while Shahar sat behind. Yoav insisted they wait until it's ready. Shahar didn’t mind. Yoav thought it was "awesome" while Shahar described it as "nice". These examples and many others demonstrate the differences between the science participation of the two children.

Differences in participation in science are often attributed to gender, race, and class (e.g., Archer et al., 2012). However, in this case, the two boys were raised by the same parents in the same everyday landscape. They were exposed to the same family habitus and were afforded the same cultural resources. Why, then, did they differ so much in patterns of participation? The answer to this question is complicated, including social and psychological components as well as historical, local and interactional ones (Bucholtz & Hall, 2005). For example, Shahar is older than Yoav, already entering adolescence, which may have been involved in how he wished to position himself in his family. Yoav, as the youngest in the family, may have been more outgoing in exhibiting exuberant excitement. In addition, perhaps Shahar was simply "naturally" less inclined towards science, cognitive activity or speaking in general? Perhaps he was more of a sports person than a science person? Indeed, Shahar practiced regularly with a basketball team and travelled around the country to compete in youth leagues. Yoav played soccer but this did not take as big a role in his activities as Shahar's basketball. Thus, Shahar's identification as a basketball player may have distanced him from science (allowing Yoav to distinguish himself as a science person). But is there more to it? Many children consistently identify with science, especially adolescents from a socio-cultural background such as Shahar's. Many children like both sports and science. Why then did Shahar, in spite of his family strong science habitus, develop an antagonist science identity?

One illustrative event: "Won't you give up your snack for the sake of science?"

To offer an additional understanding of why the two children differ so much in patterns of participation, in spite of many common conditions, I zoomed in on the moment-by-moment interaction of one illustrative event. Using micro-analysis, I explored how positioning and roles may help explain the local variation in the children's identities.

On October 10, Yoav, Shahar, Dad, and Mom took a day trip to a nature reserve. They climbed a trail up the mountain, and when they took a break Shahar asked for a snack. Mom took out a bag of chips, noticing it was puffed-up. Although this was "Shahar's snack", mom's call "look what happened to it" was taken up by Yoav, who from that moment on became a central participant in the conversation.

Mom compared the bag's current shape to its shape back at home: "at home it was all like squashed, and now at the top of the mountain it's really puffed-up like a balloon". She invited Yoav to touch it. Before Yoav responded, Dad called Mom's observation into question: "I don't think it was squashed at home", but Mom confidently urged Yoav to touch it: "Of course [it was]! I know how [it was when] I took it out of my bag before, look how strong it is here". This time Yoav accepted the invitation, calling excitedly "Yooyo, Daddy, feel here". In so doing, Yoav ruled in favor of Mom's observation, acknowledging that the bag of chips was indeed exceptionally inflated. In response, Dad removed himself from the scenario, saying he had to move out of the sun. Thus, from the very beginning of this interaction, Yoav was positioned (by himself and by his parents) as a competent, effective participant, who has the power to adjudicate between Dad and Mom. Shahar, on the other hand, was excluded. He made no effort to participate, nor did his parents or brother attempted to include him.

With the controversy resolved, both Mom and Dad adopted a teacher role, asking questions to which they knew the answers and providing explanations and illustrations, addressing Yoav and positioning him as their student. Yoav actively participated, answered and asked questions as expected – playing the role of a cooperating
student. Shahar was silent. Finally, Shahar reminded everyone that the bag was brought out for him to eat the snack.

Shahar demanded his chips: "May I eat my chips already?!" but in effect he demanded the end of the science conversation (or lesson) in which he was not a participant. Mom complied, suggesting they continue the inquiry at home where they will see that the bags are usually not as puffed-up: "Remember this and at home I’ll show you what it looks like, okay?". This time she used the plural form, addressing both children, which could have afforded Shahar's inclusion in the discussion. However, Shahar excluded himself by refusing to take part in making the bag an object of scientific inquiry, demanding "now open it". His refusal to participate or even to allow the scientific inquiry was aggravates by Dad's attempt to prevent the opening of the bag: "don't open it, let's take it home". This triggered a negotiation between Mom and Dad about the rigor of scientific evidence, with Dad insisting that to prove the bag became puffed-up with increased altitude they need to take "the same one" back home, and Mom arguing they don't: "What, so one’s puffed-up and the other isn’t?". To end this dispute, Mom turned to Shahar, asking: "Shahar, won’t you give up your snack for the sake of science?". Practically, her question put Shahar on the spot, forcing him to declare whether or not he is in favor of science. While Yoav was clearly in favor of science, begging Shahar to eat a different snack instead, Shahar refused and opened the puffed-up bag of chips.

There are multiple explanations for Shahar's behavior. However, whatever the explanation is, the result was that for that moment the family divided into two groups: the science people, including Dad, Mom and Yoav, and the science antagonist, Shahar. Thus, while in this short event, Yoav was identified as a science person, this identification also worked to identify Shahar as a science antagonist.

Analyzing the examples provided above and others, using microanalytic methods combined with ethnographic understandings, yielded the same pattern: repeated positioning, by all members of the family, of Yoav as the science person and Shahar as the science antagonist. The relative positioning of the two brothers consists of the recognition each of them received, the roles assigned to them and the roles they took on. Yoav often took on the role of a collaborative science participant, positively responding to his parents' questions or suggestions, while Shahar does not, even when the parents approach them both. Yoav was also assigned the role of a collaborative science participant by his parents, for example, when they called or addressed him only (or first), even when Shahar was also present, encouraging him to attend to an animal, phenomenon or idea (e.g., "what do you think, Yoav?"). Both these types of interactions (Yoav taking on a role or assigned a role) positioned him as a science participant while positioning Shahar as a non-participant.

Discussion
I have systematically analyzed recordings of science engagement in one science family throughout one year, showing how science in this family was infused in multiple aspects of everyday life. All family members were involved in various forms of scientific practice and content, in various settings, utilizing various tools and materials, and expressing a broad array of disciplinary emotions. However, whereas this engagement supported the development of a science person identity for one child, that of an interested and capable science participant, it alienated the other child from science. Shahar was not only failing to develop a science person identity. He developed a science antagonist identity. This did not happen in spite of the family's push towards science. Apparently, this happened because of this push.

Why then in spite of the two boys growing up in the same family, did the family habitus lead to identification with science for one boy but alienation from science for the other? I argue that repeating events of identification within family interaction can, at least to some extent, explain such differences. I follow Wortham's (2006) classroom account of repeating events of identification, in which a student was "routinely identified as a disruptive outcast" (p. 6). I suggest that routine identification through repeating interactions within the family can create a science identity trajectory of science antagonist. Whereas Wortham (2006) and others (e.g., Archer et al., 2012; Calabrese Barton et al., 2013) attribute such processes mainly to socio-historical categories, such as gender, class, and race, these categories cannot explain the differences in identity trajectories of two boys in the same family. The analysis I presented suggests that science antagonist identities may also emerge in other ways.

Family dynamics of roles and positioning are intertwined with science engagement and science identity formation. Thus, the family role in shaping science identity is not limited to affording resources, making science capital available, acting as roles models and encouraging children to engage in science. The family role includes also positioning and recognizing children as science participants and through that equally distributing access to such resources. The distribution of access to resources is not only circumscribed by socio-historical constructs, it is also dependent on subtler cumulative moment-to-moment interactions, in which members of the same family gain more or less access to resources. Thus, also in families with a rich science habitus, children can have limited opportunities to leverage available resources for the development of a science identity.
The role of the science family in shaping its members' science identity is fundamental because, as I demonstrated, everyday science engagement in such a family affords children with abundant opportunities to participate or to reject participation in doing science, in ways that are probably less prevalent in schools. Thus, it is important for parents to develop awareness of the ways positioning and roles are intertwined in everyday science engagement, and with identity formation.

Future research should look into the ways science identities develop in other families, tracking development across even larger time scales, as well as across formal and informal contexts. To enable comprehensive recording of naturally emerging family science engagement, researchers need to develop additional methodological approaches, such as combinations of first-person video recording, reflective journals, and experience sampling methods.

More broadly, this study indicates that informal science learning environments should be considered not only as facilitating the construction of positive science identities but also as leading to the development of science antagonist identities. While research has attended to such issues in school contexts, it's time to start addressing them also in informal contexts, in which people may more easily try out various roles and varied practices. While continuing to investigate and highlight how socio-historical categories work to deny individuals' (and groups') access to science, we should also look beyond these categories to expose how equal access to science is denied in other subtle sometimes less apparent ways.

References


A Focus on Contribution Towards Product and Performance in Collaborative Design

Colin Hennessy Elliott, Sarah C. Radke, and Jasmine Y. Ma
che217@nyu.edu, scr274@nyu.edu, j.ma@nyu.edu
New York University

Abstract: This paper presents a comparative case study to illustrate how resources are used for contribution in collaborative design. One case draws from video data of an urban high school robotics team which designs a robot for competition, and the other from a program where youth produce a documentary film for a public showing. Building on a view of participation as learning (Lave & Wenger, 1991), we use video-based micro-analysis to make sense of how members participate through contributions to design. We frame learning as a collaborative practice where contributions are used to make arguments with different resources. Across both cases, materials and laminations of space-times (Leander, 2002) are used as: 1) evaluative resources, and 2) knowledge constructing resources for design. In this view of learning, the final product and performance are not only end goals but used as resources in the design process through laminations of imagined future space-times.

Purpose
Collaboration is an important practice in the learning structures of Problem Based and Project Based Learning (Barron et al., 1998) as well as many other pedagogical practices in school (e.g, learning communities, Brown & Campione, 1994). Learning Sciences literature has shown that youth engage in learning through collaboration in many settings outside of school as well (e.g., The Fifth Dimension, Cole & Distributive Literacy Consortium, 2006). Further, the effectiveness of collaboration as a learning structure is dependent on the contexts and quality of interactions (Barron, 2003). Therefore, it is important to gain an understanding of how collaboration unfolds in learning settings that are not directly framed as part of schooling. This case comparison study focuses on making sense of the participation structures of collaboration, particularly how to contribute to design, for two distinct after-school activities involving secondary school youth.

In each of these cases, although we classify them as learning spaces, content learning is not typically the primary goal of the community. Rather, an end-of-semester product and its performance drive the goals of all activities for these groups; yet learning is necessary to participate in the knowledge-building process of design. The first case focuses on youth participants on an urban high school robotics team that co-produced a robot (a product) in 8 weeks, and then used the robot in competitions (a performance). The second case is set in an afterschool program called Digital Studio (all names in the paper are pseudonyms). Over the course of 14-16 weeks, youth, typically aged 16-21, developed, captured, edited, and publicly presented (a performance) a 20-minute social justice/critically engaged documentary (a product) about a social issue that impacted their lives. We classify each setting as a “semi-formal” learning space because, while occurring temporally outside of school they still exist in spaces and places that are typically considered part of school.

In both cases, being part of the community means taking action and contributing to the final product and its performance. We frame the activities portrayed in the chosen episodes as indicative of contribution, building on Lave and Wenger’s (1991) view of legitimate participation, to understand how participants contribute in collaborative interaction towards collective goals. Analyzing learning with a focus on the community provides us with the ability to make sense of contribution to design as a kind of participation. Using such a lens for analysis crafts a view of learning that privileges the shared collaborative space rather than a view of an individual’s participation trajectory within a community of practice that most analyses of learning centered on participation take. Further, it pays particular attention to the socio-historical and material contexts of contribution by members of the learning community and their differing opportunities to contribute.

In this study we analyze a set of interactions from each case, seeking to compare the different dynamics of each group engaged in collaboration as an ensemble (Ma & Hall, in press). We are interested in how these collaborative episodes unfold, what resources are used and how community members (youth and mentors/coaches/facilitators) use them to make arguments about possible design. Particularly, we are interested in how the imagined futures of the product and performance, as a lamination (Leander, 2002) of the driving goals of the design activity, are animated in the interaction and how they are taken up. The study’s guiding questions are:
What kinds of resources are used for contribution to the collaborative design activity in the process of making arguments?

How are the imagined future products and performances used as resources in interaction for the purposes of collaborative design?

Collaboration-as-learning: A Collective Approach

Taking a sociocultural view of learning and development informed by Lave and Wenger’s (1991) conception of learning as a progression of participation we view the design practice of both settings in this study as instances where participants of a community learn and develop by collaboratively engaging in creative activity. Further we take up the lens of “collaboration-as-learning” (Enyedy & Stevens, 2014) where the analytical unit of interest is the collective group engaging in, in this case, a design task with a focus on describing and making sense of how collaboration is organized and enacted.

In the cases of this study, the learning environment is framed around activities of collaborative design by a community of practice (Lave & Wenger, 1991) with the goal of creating a product to be used in performance. Therefore, with an analytic focus on the collective work of collaborative design, the way contributions to the design are brought to bear is important in making sense of the participation structures. Here we focus on contribution as a particular type of participation, specifically how contribution happens including what resources are needed and used to collaborate.

Contributing to the design of a product and performance like in these two cases requires access to resources for contribution which can be brought to bear on the interactions of collaborative work. Influenced by Latour’s (2005) view of social practice in scientific work as ongoing assembling of associations of material and human experiences, we treat the work of these communities of practice as authentic authoring of activity where associations are assembled to contribute to design. The remaking of associations takes the form of resources used for contribution; these resources are animated interpretations that participants bring, through interaction, to the collaborative activity to make an argument. They are derived directly from the material experience at hand in the interaction, from a remaking of past space-time, or imagining a future space-time (Leander, 2002). Building on Leander we call “laminations” the co-creations of space-time representations (chronotopes), or instances where members of an interaction take up, add to, and utilize social representations of past, present, or future space-times.

Data and methods

Data comes from two distinct studies of semi-formal learning. In each study data includes video and audio recordings, ethnographic field notes, and images. Hennessy Elliott collected the data of the Robotics Team analyzed in Case 1. Data from Case 2 came from a larger study of the Digital Studio by Radke and Ma. Analytically, we focus on interactions where youth contribution to collaborative design of the product and performance is the driving activity. This focus structured our search for and selection of episodes for analysis to those in which the facilitators were either not physically present or were not directly guiding the activity. The episode for the Robotics Team in Case 1 occurred just after the competition rules for that year were released. It portrays a portion of the team collaborating with two coaches and a mentor on crafting robot design ideas and competition strategy. The episode from Digital Studio in Case 2 also occurred early in the design process (in this case, of the documentary) and shows a subgroup of youth building a topic idea to pitch to the rest of the group. Methods of multimodal and interaction analysis (Goodwin, 2010; Jordan & Henderson, 1995) were used to examine instances of contribution during design activity.

Case 1: Robotics Team

The Robotics Team’s home is in the basement mechanics shop of Engineering High School (EHS) in a midsized northeastern city. EHS is a technology and engineering focused high school that serves students from all over the city, the large majority of whom are Black and or Latinx. The team has participated for over 20 years in annual robotics competitions which are re-imagined and remade every year by an international governing organization. The particular episode for this study was chosen because it occurs shortly after the year’s competition rules were released. It portrays a moment where different members of the team collaborated on 1) developing ideas for this future robot, and 2) developing a better understanding the rules of the new competition style (see Table 1 for a summary of all participants in the episode). The coach, JF, had a youth participant print out the competition manual and put it into a three-ring binder for access during discussions like the one analyzed here. In this episode, Matt, the college student mentor, holds the manual near the table that all of the participants are crowded around.
Table 1: Participant list robotics club

<table>
<thead>
<tr>
<th>Name</th>
<th>Major Position/Duties</th>
<th>Grade</th>
<th>Year # on Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jav</td>
<td>Captain, builder</td>
<td>11th</td>
<td>3rd</td>
</tr>
<tr>
<td>Spock</td>
<td>Builder</td>
<td>10th</td>
<td>2nd</td>
</tr>
<tr>
<td>Harrison</td>
<td>Support for scouting and design</td>
<td>12th</td>
<td>1st</td>
</tr>
<tr>
<td>DV</td>
<td>Math Teacher and Coach</td>
<td></td>
<td>1st year as coach</td>
</tr>
<tr>
<td>JF</td>
<td>Coach and Engineering Teacher</td>
<td>Teacher</td>
<td>4th year as Coach</td>
</tr>
<tr>
<td>Matt</td>
<td>Mentor/ Driving coach</td>
<td>Alumnus, current undergraduate</td>
<td>2nd year as mentor</td>
</tr>
</tbody>
</table>

Three youth members of the Robotics Team collaborate with two coaches and a college mentor (See Table 1) to talk through the design of a robot and the new rules of competition that would influence both in-game strategy and how the robot should be designed. This analysis focuses on the resources brought to bear on this set of interactions over a five-minute period.

Multiple times in this episode, members use references to previous competitions in order to gain traction with their own ideas for how the future robot should be designed. For example, responding to JF’s suggestion to build a conveyor belt in order to gather competition balls that need to be shot into a certain goal, Spock attempts to contribute to the design discussion by sharing an idea in terms of another robot he saw at a competition the previous year:

**JF:** No matter what the balls have to be picked up. Ahh conveyer belt won't be a problem to do. ((looks down to the table))

**Spock:** Are we gonna do it, uhh, Remember last year's team. The giant one with the ((gestures up and points with right hand, index finger; looks away)) like the … it was... Basically, last year there was a um, a robot that ((gestures with arm as grabbing something)) grabbed the ball and it was like a conveyer belt ((spins arm in a circle)) and it just went straight into the shooter and they just kept like rotating the whole time...

While he struggles to get his ideas out, Spock’s continued gestures fortify his attempt to describe a past competition that they should learn from. Here, the experience of a past time-space at a previous competition is a resource for contributing to the design discussion, linking to JF’s idea that this year’s robot could have a conveyor belt for picking up balls. This is an example of laminating a past space-time into the present as a resource to build on the collective knowledge of a design. Once brought into the discussion it is picked up and responded to, particularly by JF:

**JF:** Yeah but that. Remember how tall that was.

**Spock:** Yeah but it. Can we do some[thing like that]

**JF:** [And there were only]. They were only doing it with one ball.

**Spock:** Yeah but I mean is, can we do something like that but on a larger scale like make it wider ((gestures out with both hands, a widening gesture))

In this exchange, JF rebukes Spock’s contribution with another frame of the previous year’s competition, “remember…” only one large ball was picked up and shot; whereas in the current competitions multiple smaller balls needed to be picked up and shot at a goal to score points. The “but…remember” sequence frames the exchange as an evaluative resource that references the same space-time. Here, while referring to the same historical time-space as Spock, this is also an example of bringing the resource of the competition rules to this collaborative design process in order to make an argument against basing their design off of the previous robot Spock refers to. Later, JF, as the mechanically inclined coach, clarifies that a design suggested by Spock and Jav was an “overly complex design” that would have too many spots of possibly not working, using mechanical knowledge to laminate an imagined future space-time. In this comment JF brings the proposed robot design into talk by describing a space-time where something could go wrong if the robot were designed that way.
Besides the mention above by JF, the competition rules, as a resource, are brought into the collaborative discussion into the interaction in two different ways. First, they are recruited as a reference to physical manipulations of the game pieces built to scale out of wood in the shop around them. Second, they reference the competition game manual printed and placed in a binder which sits in front of Matt in this specific interaction. Both examples mediate the rules of the competition through physical artifacts that can be referenced.

At one point in the collaboration, the group begins to construct a strategy for how the robot should move around the competition field, in addition to constructing a working design of the robot itself. In the example below, Matt (the mentor) is seated in front of the game manual which is flipped to the page with a figure that shows the layout of the competition field with dimensions and labels:

Jav: Everything's fair game in this
Matt: You have to go here ((points end of screwdriver to field map)) and then ((drags screwdriver to a new spot)) right here ((moves screwdriver back and forth)) diagonal
Spock: I mean I'd guard our neutral zone ((points to the field map)) or retrieval zone.
Matt: [Cause if you] go down here you run that risk as well. ((Points, with screwdriver to the field)) Remember there's one here. ((moves screwdriver to another point on field map)) And there's one here.
Spock: Que--- I mean ((points to the field map)) uh…

Competition strategy has large implications for design choices here. In this set of utterances, Matt uses this representation of the field, or field map, to argue his view of what their strategy should be in the competition. The game manual animates the rules of the competition and structures its vocabulary (“neutral zone” and “retrieval zone”) for Matt, Spock, and those looking on. It also directly mediates the lamination of a future time-space where an imagined robot moves through a competition (“you have to go here”). As Matt points with the screwdriver to the map of the competition field, he merely has to bring the imagination into the future while the map in the manual does the work of constructing the spatial relationships for other participants. The screwdriver he touches the page with becomes the future robot. Spock takes this up pointing to the map and also speaking in future tense “I’d guard…” These resources, the map in the competition manual and space-time representation, are used to build the collective knowledge of the rules (“you have to…”) and competent strategy in future performance (“you run the risk of”)

The use of the physical game manual is picked up by JF, and DV, all referring to the representation of the field in order to bring the future competition into the design conversation, as seen in the excerpt below.

JF: Is there a danger zone in front of the, ((leans in to the binder)). Oh there it is. Yeah I see it
Matt: Right here
DV: So Andy was talking about that pattern ((Spins finger in a vertical circle)) (0.4s) remember, the board he had that loop pattern? … So I think that loop pattern ((Spins hands in loop)) will keep us away ((points to the field map)) (0.2s) from the penalty spot.

In these turns at talk, the game manual becomes an integral resource for the collaborative design process, even mediating other space-time resources. The two excerpts above show that the manual is a materialization of competition rules, which is taken up to laminate the future performance of the robot to make arguments about how the robot should move, and how it should be designed. As these contributions are brought to bear on the interaction the learning community collaboratively negotiates their implications and legitimacy.

Case 2: Digital Studio

The Digital Studio is a non-profit youth media organization focused on engaging underserved youth populations in documentary film production and civic engagement. The program is offered for school credit or as a paid internship and emphasizes community building and cooperative problem solving. Over the course of one school
semester, interns work together to develop, capture, edit, and publicly present a 20-minute documentary. Starting with students’ own questions and their own lived experiences, these documentary inquiry projects look to challenge social issues and institutions. Digital Studio youth are supported by one adult facilitator, but ultimately, they were responsible for all the footage collection and editing, as well as for developing and realizing a strong story line (called the line of inquiry).

The episode below occurred early in the semester as the interns worked to choose the topic of their documentary. We meet six male youth as they prepare for their pitch presentation where they will attempt to convince the whole group to select their production idea. The episode was chosen because it was a moment of facilitator-free collaborative negotiation in an effort to 1) formalize their answers the Pitch Packet questions, and 2) develop or share understandings of the critical components of a documentary film. Table 2 provides names and other pertinent background information for the participants in this episode.

### Table 2: Participant list Digital Studio

<table>
<thead>
<tr>
<th>Name</th>
<th>Role</th>
<th>Age</th>
<th>Grade</th>
<th>Other pertinent information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benjamin</td>
<td>Intern</td>
<td>18</td>
<td>11th</td>
<td>(1)</td>
</tr>
<tr>
<td>Carl</td>
<td>Intern</td>
<td>17</td>
<td>10th</td>
<td>(2); First language is Spanish</td>
</tr>
<tr>
<td>Elliot</td>
<td>Intern</td>
<td>16</td>
<td>10th</td>
<td>(2); First language is English</td>
</tr>
<tr>
<td>Ian</td>
<td>Intern</td>
<td>17</td>
<td>10th</td>
<td>Nominated to be the scribe</td>
</tr>
<tr>
<td>Jorge</td>
<td>Intern</td>
<td>18</td>
<td>11th</td>
<td>(1)</td>
</tr>
<tr>
<td>Nicholas</td>
<td>Intern</td>
<td>18</td>
<td>12th</td>
<td>(1)</td>
</tr>
<tr>
<td>Hank</td>
<td>DS Facilitator</td>
<td>N/A</td>
<td>N/A</td>
<td>(1); Second year as a DS Facilitator</td>
</tr>
</tbody>
</table>

(1) Does not appear in transcript excerpts but is part of the data analyzed for this paper.
(2) Recent immigrant to the United States

The small group collaborates to finalize their documentary idea and complete their Pitch Packet. This packet, created by the facilitator, Hank, is a collection of questions they must be able to answer during their formal pitch. The packet’s questions are divided into the following five categories: Line of Inquiry, Personal Profile, Relevance, Professional Interviews, and Closing Argument. They will have to present a formal pitch later this same day and have been researching and developing their idea for the past three days. During the formal pitch, the packet they are finishing now will be available to them as a reference for their talking points. Similar to the Robotics Club, resources are called upon through talk and gesture with reference to experiential time-spaces, past and future, and material resources physically available in the room. These are used to contribute to the collaborative design for both evaluation and knowledge building purposes. In the following excerpt, their work to finalize their pitch turns to a discussion of one aspect of the documentary, the personal profile.

**Carl:** Oh, and also for example history profile ((Gestures towards Nicholas)).

**Ian:** What? ((Holding the Pitch Packet))

**Carl:** The history of? profile.

**Ian:** Personal profile?=

**Carl:** =Uh huh?

**Ian:** Like, every single video that you watch uh, is is connected to uh personal profile. So it’s, i-if you don’t have a personal profile you cannot, you cannot start working on your documentary.

In this excerpt Ian, who is in charge of recording their ideas in the pitch packet, leads the discussion. Carl’s first utterance was originally in response to the question, “How will this be different from other Digital Studio films?” However, Ian takes it up as a content question. This is evidenced in his reply of what a personal profile is and why it is important to a documentary. Carl rejects Ian’s interpretation of his idea as a request for information.

**Carl:** No but I say, no but I, no, I means that they history, in the history, always different for each people uh the history of the don’t trust the internet like that.
Ian: Well::: (Still holding the Pitch Packet)) I didn’t get your question I’m sorry but like I want you to elaborate a little bit for me, [like explain it.

Carl: [Yes. For example - for example, not to trust the internet because I ha:::ve one day I have to computer it? ((make typing motion)) that’s why - I was::: five years old, no? ((gestures with hand to the height of a five year old)) And I put uh, computers like that ((typing motion)) and appear a virus in my computer like that..and I don’t trust the internet. Another person for example, hack or like that or um, anything else, anything like that um::: don’t trust the internet because eh, security of the pages is low.

Ian: Okay, I got, I got what you said, but is that, is that a real personal story? Like, it-it must be real. It-it must be something that happened to you already. I-you cannot make from your own. That’s what I was tryin’ to do. When - when I spoke to Hank, told me like, this should be a real his-story.

Carl’s example seems to clarify for Ian that he understands what a personal profile is. For Ian this leads to a new issue of authenticity. We see Hank animated through Ian’s talk when he refers to a past interaction (“Hank told me...”). Carl never replies to Ian’s authenticity question and the group moves on when Carl admits that he is, “sleepy” and “[doesn’t] even know what [he’s] talking about.” Because of his responsibility to record their ideas, Ian has been positioned as the leader. His uptake of this role is evidenced by his multiple returns to the pitch packet from which he produces lines of direct questioning, one of which he poses to Carl in the excerpt above. Furthermore, he animates lessons from Hank (“it must be real”) to clarify what counts as a personal story and why it is relevant to documentary making. Finally, he uses a past experience with Hank (“when I spoke to Hank”) to evaluate their answers and work to negotiate the final record. Across the excerpt the pitch packet acts as a mediating artifact, connecting the lessons on documentary making Hank has taught and their current collaborative work efforts.

The pitch packet also becomes useful in elucidating a historical framing which the youth take up in developing and evaluating their answers. This first occurs when the youth work to answer the question, “How will this be different from other Digital Studio films?”

When Nicholas begins to say their topic has never been covered, Elliot interjects, relying on his knowledge of past production topics. Another student, Ian, recalls that the Digital Studio has produced over 190 documentaries so, “even Hank haven’t watched uh the videos, like all the videos, there might be one.” In this utterance grouping the youth work collaboratively to evaluate the strength of their originality argument. Ian animates the then-absent facilitator, Hank, in reminding them of the multitude of completed projects. This could be taken as an evaluation of Nicholas’s first comment and support of Elliot’s warning that they “don’t want to make assumptions,” about uniqueness. This historical framing then becomes crucial in formalizing their final argument. The three youth decide to skirt the absoluteness of Nicholas’s original statement and change their answer to say, “It’s an unexplored topic.”

Across the episode learners collaboratively produce the documentary and position both the pitch packet and the facilitator as resources in their work to develop and finalize their group’s documentary idea. In these excerpts, the pitch packet becomes an integral resource for the collaborative design process. It laminates their shared lessons from Hank about documentary making and the future completed film as resources for completing the final draft of their pitch. The interns bring three different kinds of resources to this collaborative exercise: 1) shared historical framing, 2) the pitch packet, and 3) Hank’s ideas and lessons. These resources get recruited by the youth producers in the collaborative process of negotiating and evaluating ideas as well as in support of knowledge-building.

Discussion
Collaborative design work is driven by the ways resources are brought to bear on the interaction in each of these settings. As Barron (2003) argues, focusing on the group as the analytic unit allows us to analyze the resources they use, the “types of contributions they make and how they are taken up or not,” (p. 311). Learners interact in authentic practice of designing and do the work to position each other as active members of the “epistemic community,” (Kim, et al., 2015). In the theoretical view we take, this positioning in the community is facilitated by learners’ access to usable resources, crafting their context for contribution and the ways contributions are legitimated by the group. Across both cases these resources are used as: 1) evaluative resources, or resources to legitimize an emerging part of the design, and 2) knowledge constructing resources, or resources to further construct local knowledge needed for the design or as part of the design. In each analysis it proves difficult to
separate product from performance in what future collaborators are designing. Therefore, both cases take on future oriented analyses that do not directly distinguish the two.

Above we identified three resources used in interaction in each case. Each resource was used, at times, in both evaluative and knowledge constructing contexts. These two lists of three, while somewhat different, map quite well onto each other. First, previous competition experience, a resource identified from the Robotics Team case, matches well with Digital Studios’s shared historical framing. Each resource indicates the agentic use of the institutional historical contexts. This points directly to the importance of historical framing and experience in how youth can contribute to design work. Second, both the competitive game manual (Robotics Team) and the pitch packet (Digital Studio) are physical material resources that facilitate the interaction and are used to structure the design process including creations of imagined future space-times. Third, Hank’s ideas and lessons (Digital Studio), which youth take up, can be construed as the mechanical (technical) knowledge of making a documentary. JF and others clearly use mechanical knowledge of robot construction to design and evaluate possible design.

Youth interpretations of past space-times were used in both cases to help construct a vision of a future design. On the Robotics team, previous experience at a past competition was repeatedly used as a resource through lamination of past time-spaces. In the Digital Studio, youth reimagined the past program documentaries to argue for the originality of the idea they were proposing. Both are examples of how this type of contribution resource requires knowledge of or experience with past iterations of the type of product they were aiming to design. In the Robotics example, the learner had the previous experience and therefore was able to bring that to bear on the interaction though a lamination of a past space-time while the learner in the Digital Studio example needed to speak to a historical context by referring to previous time-spaces where the facilitator discussed yet another previous time-space, limiting the possible lamination. This is a secondary layer of space-times for reference.

Kim and colleagues (2015) describe integrating the design process into classroom learning as an “emerging learning process” where learners define the learning trajectory. This lens can also be taken in these semi-formal learning spaces. Ways to contribute are continually re-created depending on both the resources at hand for youth, how previous contributions have been legitimated, and the collaborative interactions that come before it, laying the groundwork for what is needed to be learned in order design and create the artifact. This required us to pay attention to the ways that future products, and in these two cases their performance, were brought to the design process as both evaluative resources and knowledge construction resources. In both cases, the groups interacted, constructing imagined products with past and future space-times, and articulating understandings of the goals of future performance to make arguments.

Future space-time representations are particularly important in design work for a product and a performance. Participants frequently try to make arguments and build consensus for a design choice by animating their imagining of a future product and/or their performance through talk and gesture. Learner contribution to the design process in both cases required directly constructing and sharing an imagined future of the final product and the final performance of the learning community. Therefore, the product and performance become part of the design process, instead of simply being end goals. Learners, and mentors in the case of the robotics team, used resources, including imagined and interpreted space-times, physical objects, and technical knowhow, to bring these final goals into interaction in the design process. In both cases these resources were used to evaluate a current design idea and to build collective knowledge about the rules for the products and their performances and the design itself. While the pitch packet and the game manual, themselves, were designed to do this work for learners in collaboration, many other time-spaces and objects were used to further construct design ideas. These case analyses illustrate the nuanced ways youth contributing to design discussions bring together constructions of varied space-times, including interpreted past and imagined future, as resources in conjunction with the material resources at hand, i.e. the pitch packet or the game manual.

Therefore, in collaborative work for learners in semi-formal learning spaces, it is important to recognize the different resources (material and temporal-spatial) that learners have access to, how contributions are collectively taken up, and to be intentional about how they are facilitated. It is that access, or perceived access in conjunction with the ways the group takes up learner contributions which gives learners the agency to contribute to the design and, in turn, the construction of the learning environment.

**Significance**

The analysis showed the ways in which collaboration is guided by the youths’ knowledge and imaginings of their end goals in each setting—both the physical products and the final performances. Hmelo-Silver writes, “the goal of becoming a good collaborator and the process of learning collaboratively are often woven together,” (2004, p.214). With this view of the learning community, it becomes clear that differentiated access to resources, both material and spatio-temporal, structure the types of contribution learners can collaboratively make to a design process. Further, the ways in which contributions are valued, evaluated, and taken up play a large role in later
access to resources and contribution. Therefore, designing and facilitating these types of learning spaces which include a focus on a final product and/or performance requires paying attention to learners’ previous experiences, their developing imagined final product and performance, and their material contexts for design. This is also important for crafting more formal learning spaces with project-based and problem-based learning pedagogies.

References

Acknowledgments
We are thankful to our colleagues in the New York University Interaction Analysis Lab who participated in data sessions. Their analytic expertise greatly assisted this research. We are also grateful to the insightful reviewers who provided comments that pushed our thinking.
Exploring Multimodal Scaffolds Supporting Middle School Students’ Construction of Causal-Mechanistic Scientific Explanations

Nitasha Mathayas and David E. Brown
mathaya2@illinois.edu, debrown@illinois.edu
University of Illinois at Urbana-Champaign

Abstract: Computer simulations help students visualize abstract scientific concepts. Yet their effectiveness depends upon the kinds of supports students receive. In this study, we explore the verbal and gestural supports we provided to students as they viewed a computer model of molecular conduction and constructed causal-mechanistic explanations for heat transfer. Using a complex systems and embodied approach to students’ conceptions, we conducted a cross-comparison case study of four students’ experiences with the simulation. By comparing students with richer scaffolding interactions to those with weaker scaffolds, we found that directing student attention towards causal elements in the simulation along with requesting them to gesture about the function of those elements helped students develop sophisticated explanations. We describe these cases and reveal ways these scaffolds can and cannot be enacted with the simulation. Implications for instruction and future work are described.

Introduction
In the United States, constructing causal explanations is one of the eight essential practices listed by the Next Generation Science Standards (NGSS Lead States, 2013). Computer simulations can help students to visualize hidden and unobservable elements underlying several physical phenomena, such as the motion of air molecules causing pressure or the function of light rays in the working of mirrors, the formation of shadows, and in the occurrence of seasons. Moreover, with technological advancements providing immersive experiences for learning (Lindgren & Johnson-Glenberg, 2013), students can partake in abstract scientific concepts in embodied ways where they can physically enact the behaviors of unobservable elements (e.g. Johnson-Glenberg, Birchfield, Tolentino, & Koziupa, 2014). However, the effectiveness of a simulation depends upon several factors such as the kinds of support provided to learners, the design features of the simulation, and the role of the teacher in guiding students through the simulation (Smetana & Bell, 2012). Hence while immersive technologies continue to advance into the educational sphere, understanding the role of facilitation with and without technology is still a priority. In this sense, our work is aligned with the goals of this conference where we are developing gesture-augmented computer simulations that support the construction of causal-mechanistic explanations (Braaten & Windschitl, 2011), while also examining how we can facilitate conceptual engagement using these simulations.

Our research is part of a design-based research project developing gesture-enhanced simulations about heat transfer, air pressure, and the causes of the seasons over the course of four years. In the first year, we explored how students engaged with a mouse-controlled computer simulation that depicted a molecular model of heat transfer. We observed how students naturally gestured during their explanations and we explored different ways we scaffolded (Wood, Bruner, & Ross, 1976) the development of these explanations. The analysis of this data was then used to develop gestures that supported students’ scientific explanations in the successive years where a gesture-enhanced simulation was used. In this paper, we focus on different ways we supported student interactions with the mouse-controlled computer simulation, and we examine how our interactions with students helped them use relevant ideas from the simulation to construct scientific explanations of heat transfer. Our research questions are: (1) Did students’ explanations improve after they interacted with the computer simulation? And (2) How does multimodal (i.e. verbal and gestural) scaffolding influence student construction of causal-mechanistic explanations?

Theoretical orientation

Students’ conceptions in science
Brown (2014) has proposed that knowledge is a complex system composed of a large number of knowledge elements that are dynamically interacting with one another. Conceptions are structures that emerge from the dynamics of this system, and stable conceptions are robust structures that persist through changes in the system. There is also mounting evidence that scientific concepts are grounded in embodied schemas (Amin, Smith, & Wiser, 2014). Our sensorimotor perceptions inform and intuitively guide conceptual development. This
perspective implies that learning is a dynamic, evolutionary and nonlinear process where conceptual structures are embedded in larger knowledge systems involving both perceptual and mental states.

Embodied cognition
Research in cognitive science, MRI scanning, and psychology has shown that human cognition has deep roots in sensorimotor processing (Wilson, 2002). Barsalou (2008) described cognition as grounded in “simulations, situated action, and on occasion, bodily states” (Barsalou, 2008, p. 619), in which the brain’s modal system for perception (e.g. vision, audition), action (e.g. movement) and emotion are involved in the formation of a mental simulation. In addition, Smith (2005) described cognition as a complex dynamic system where “intelligence emerges in the interaction of an organism with an environment and as a result of sensory motor activity” (Smith, 2005, p. 278). Consequently, conceptions are hypothetical entities representing stable states of a dynamic system. These stabilities emerge from the dynamic coupling of the body with the environment, and stable states gain resilience when corporal modalities are synergistically linked with each other and with the physical event experienced at that moment in time.

Applying ideas from both domains, we think of students’ explanations as dynamic constructions in which prior knowledge and sensorimotor experiences contribute to the dynamic formation of a conception. For instructors supporting students during their active construction of mechanistic explanations, scaffolding that addresses embodied resources and integrates multimodal representations can be particularly fruitful in fostering stable scientific conceptions. However, this perspective does not explain what scaffolds look like in instruction, or how, and when they should be implemented. Taking Wood, Bruner, and Ross’s (1976) notions of scaffolding functions as a guiding principle for this study, we conducted a preliminary investigation about the role that scaffolding played when we facilitated students’ use of gestures and a computer simulation while constructing causal-mechanistic explanations. We identify these scaffolding moves in this paper and describe the multimodal nature of these scaffolds.

Methods
For this study, we conducted a collective case study (Creswell, 2013) using select student interactions with the computer simulation. In this method, multiple cases are examined to better understand the phenomenon or event. While there is redundant information across the cases, this redundancy is either used to draw more compelling arguments (Barone, 2011), or it is used to investigate the nuances between cases that leads to a better understanding of the phenomenon. We applied the latter meaning in our work, where we examined four students to better understand how scaffolding interactions in some cases led to complete causal-mechanistic explanations, while other interactions led to partial explanations. In the following subsections, we describe the context of the study and how our cases were selected.

Context
In the first year of the larger project, the team of researchers interviewed 36 middle school students (grades 6 to 8) from the surrounding public schools of a large Midwestern University in USA about their understanding of three scientific phenomena: thermal conduction, air pressure, and the cause of seasons. Of these, 24 students (15 boys and 9 girls) were interviewed on the topic of thermal conduction. The interviews were designed to be semi-structured with three general phases: before, during, and after viewing a computer simulation depicting a molecular model of thermal conduction. Following are details of the conductive heat transfer interviews.

Interview structure
Phase 1
The interviewer introduced a silver spoon partially immersed in a cup of hot water and let the student experience heat transferring along the handle of the spoon. The student was then prompted for an explanation of how the handle of the spoon got warm even though it was not in contact with the hot water. In cases where students were not thinking of a molecular model of the spoon, the interviewer suggested that they consider a zoomed-in view of the spoon. Hence, the first phase of the interview involved exploring students’ initial explanatory ideas of heat transfer.

Phase 2
The interviewer showed students a simulation depicting two blocks of molecules vibrating at different temperatures (Figure 1a). The purpose of showing this was to prompt students to think in terms of dynamic molecular structures of solids at different temperatures, and to understand that molecules vibrate more at higher temperatures. The second part of the simulation depicted these blocks connected to each other through a bar of
molecules (Figure 1b). This simulation started out with the left block of molecules at a higher temperature than the right block, but when played, vibrating molecules on the left side collided with adjacent molecules in the bar prompting them to collide into their neighbors causing a chain reaction of molecular collisions to spread to the right block of molecules. This made the right block of molecules vibrate more making the temperature gauge on the right rise until the system achieved equilibrium. In this phase, the interviewer was prepared with prompts to guide the development of causal-mechanistic explanations. These prompts involved helping students connect representations of temperature with molecules, asking for predictions before the simulation was played, prompting students to use their hands to represent the movement of molecules, and making connections between the simulation and the spoon. In this exploratory phase, the interviewers added other forms of questions that they thought would help students. These explorations are included in the analysis of this study.

![Figure 1.](image)

**Phase 3**

The interviewer closed the simulations and asked again how the handle of the spoon got warm. After students attempted their first explanation, the interviewer explored different prompts to help students recall important features from the simulation including gestures they had previously used, and then had them reexplain the phenomenon using their new tools.

**Analysis**

Each interview was video recorded and edited to include the screen capture of the computer simulation used in the interview. These edited videos were then transcribed and used for analysis. First, we segmented each student interview based upon scaffolding prompts included in our interview protocol. These moves are: asking for predictions (IP), testing the phenomenon (TRY), suggesting molecules (MG), requesting a drawing (DRAW), showing simulation Figure 1a. (S1), showing simulation Figure 1b. (S2), requesting for an explanation post-simulation (Post-Sim), requesting to show what you mean (SHOW), requesting to show how the phenomenon occurs (SHOW:X), using hands to represent a part (SHOW:PART) of the phenomenon, reifying student (SHOW:ST) gesture, and prompting student to use the interviewer’s (SHOW:INT) gesture. When the interviewer employed one of these scaffolds, we marked this time in the transcript as the start of a new segment. While most of the scaffolds were used sequentially in each student interview, some scaffolds were more flexible (such as SHOW) and were used at different times, while some (such as DRAW) were not used at all. Thus, each student’s interview is composed of the three phases described previously, but is a unique sequence of segments within the phases.

Next, we constructed a target causal-mechanistic explanation of heat transferring through the spoon, parsed this explanation into six main explanatory elements (Mathayas, Brown, & Lindgren, 2016), and then coded interviews for the presence of these elements. Interrater agreement for this coding was 91.8 percent. Finally, we organized the codes into a scoring scheme that emphasized increasing causal power of an explanation using a combination of the elements. The description of each score is in Table 1. A Wilcoxon’s signed ranks test was applied to check for significant differences in scores before and after the intervention. These scores were also traced over the course of a student’s interview to create charts depicting their evolving explanations. By the end of this process, we had 23 Explanatory Expression Charts that depicted each student’s evolving explanations about heat transfer in a spoon (one student’s interview was incomplete and was left out). Based on the results from this analysis, we selected cases for a closer inspection of scaffolding described next.

<table>
<thead>
<tr>
<th>Score</th>
<th>Level of target explanation the student communicated</th>
</tr>
</thead>
</table>

Table 1. Scoring scheme
The spoon has molecules

Molecules of the spoon are in a rigid-structure and are position-dynamic

All of the above and molecules can interact and influence one another

All of the above and molecules wiggle more at a higher temperature

All of the above and faster wiggling molecules collide with slower wiggling molecules making them move more.

All of the above explained as a chain reaction along the length of the spoon

Findings
The Wilcoxon’s signed ranks test revealed that students’ Post-Simulation Scores were significantly higher than Pre-Simulation Scores ($Z = 171, p < 0.001, W = 0.783$). We also conducted this test to check for the effect of the simulation and found that During Simulation Scores were significantly higher than Pre-Simulation Scores ($Z = 185.5, p < 0.001, W = 0.661$). Therefore, in response to our first research question, students’ causal mechanistic explanations improved after using the simulation. However, we noticed that half the students obtained a full score in the post-simulation phase, while the other half reached a score of four or below. Therefore, to respond to our second research question, we selected four students that were representative of this division. On taking a closer look at student scores, we noticed that there were similar forms of unplanned moves that interviewers used that were not captured in the charts, but these questions were asked in different ways for these four students. Thus, these students were selected for a cross case comparison study driven by the theoretical orientation of conceptions as dynamically emergent structures with embodied sources. The student transcript, drawing, video recording, coding, and graphs were all used to analyze the interactions.

We now report on our analysis of Andrew, Eric, Irene, and John (all names are pseudonyms). Andrew and Irene gave a full causal mechanistic explanation by the end of their interviews, whereas Eric and John gave partial mechanistic explanations. We begin with Andrew who had the richest scaffolding with the interviewer followed by Eric, who had the least scaffolding of the four. We then visit Irene, who had a variation of Andrew’s interaction, and then John who had a lesser level of scaffolding compared to Irene. As we move through each focal student in this section, we begin with a brief description of each student’s growth on the charts. The focal student description will introduce their explanations and then delve into their responses to interviewer prompts while interacting with the simulation. Each focal student description will conclude with a summary of the main scaffolds used in that interaction.

Andrew – Full explanation, rich and varied intervention
Figure 2a represents Andrew’s evolving explanations while viewing the simulation. While he was scored low before he watched the simulation, there is considerable growth depicted in the graph around when he watched the second half of the simulation (Figure 1b). Before viewing the simulation, Andrew attributed heat transfer to a substance flowing through the spoon. He expressed this idea in a drawing where heat molecules travelled over the length of the spoon. However, when he watched the simulation, he expressed this notion of travelling again although the molecules in the simulation do not physically travel. The interviewer intervened at this point by replaying the second half of the simulation with a new question. An excerpt is shared below. These transcripts are partially edited to include relevant interactions, and the speaker’s gestures are included in parenthesis.

**Interviewer:** Now remember that scientists think of temperature as how fast the molecules wiggle (raises both hands as fists and wiggles them), and so thinking about just the molecules themselves wiggling, how does that fast wiggling over here, on the left, make these molecules (points to molecules on the right of the simulation) wiggle fast?

**Andrew:** It's almost like, it's like a chain reaction. So, like the wiggling there (points to the hotter block of molecules on the left of the simulation). They're wiggling, they're bumping into each other, which makes them bump into that (points to the connecting bar) and travels all the way over to the handle side. (uses right hand to chop the air three times while moving to his right).

**Interviewer:** Okay. Can you show how that works?

**Andrew:** Um, let me think about it, so… I don’t know how to show it.

**Interviewer:** Can you maybe use your hands to represent the molecules and show how the molecules on the left make the molecules on the right wiggle faster?

**Andrew:** So, what's happening is this hand is the molecules on the left (opens left hand up). and this is just going to be the spoon part (opens the right hand up next to the left
with finger wiggling). So, they bump into each other (bumps both hands together), so this side (indicates left hand) is just bumping, they're all bumping into each other (wiggles fingers and touches them) and then since the spoon part is right there (indicates right hand) it bumps into those molecules here (bumps left hand’s fingers into the right), which then if I move over (shifts both hands a little to his right), this (left hand) becomes the middle and that becomes the left (right hand) bumps into that side, making them keep wiggling (circles index fingers around each other).

In figure 2a this interaction occurred from where Andrew scored a 6 at S2 up to SHOW:PART. On examining this interaction from an instructional perspective, we see the interviewer making multiple pedagogical moves. First, he directly instructed Andrew about the relationship between molecular movement and temperature. He referred to the scientist’s notion of temperature to give his statement more authority. However, this instruction was meant to remind Andrew of information they had already established at S1, thus we do not consider this move as didactic, but a way to return Andrew’s conceptual trajectory towards target concepts he established during S1. The second move was to refocus Andrew’s attention to the wiggling of a single molecule. The interviewer achieved this using both speech and gestures, and thus redirected Andrew away from flow-like mechanisms to particulate mechanisms. Third, we see a causal question targeting molecules and their agency in effecting movement in other molecules. By asking “how does the wiggling on the left, make the molecules on the right wiggle more?” we see the interviewer leveraging the current focus on wiggling molecules into thinking about causes and effects. Andrew responded by using the simulation to build a causal explanation. However, this interaction does not end there, the interviewer executed two more moves related to the use of gestures. The first request was to use gestures generally to describe a chain reaction, but with Andrew expressing his struggle, the interviewer then grounded the question in the motion of molecules by specifying that Andrew gesture about the motion of molecules. This appeared to be fruitful for Andrew, who then described a chain reaction of collisions in both speech and gestures. In terms of our theoretical lens, we think this move helped reinforce Andrew’s current thinking beyond what would have occurred had he merely stated it. Since he used both speech and gestures synergistically to describe the molecular collisions, we think that the interviewer’s move was perfectly timed to help synthesize the new perspective Andrew had taken.

Summarizing this interaction, we see five scaffolds occurring in this interaction which are: 1) providing direct instruction about mutually established facts, 2) refocusing student attention to the movement of molecules, 3) causal questioning, 4) request to gesture to explain, and 5) request to represent molecules with hands to explain. We now consider these moves in the interactions that other students experienced.

**Figure 2.** (a) Andrew’s scores in response to interviewer scaffolding. (b) Eric’s scores in response to interviewer scaffolding.

**Eric- Partial explanation, minimal intervention**

Eric’s scores while watching the computer simulation (Figure 2b) appear unchanged across his interview. He expressed a molecular description of heat before watching the simulation, but he was scored for a partial explanation because he did not attribute causation to molecular collisions but to friction. As he watched the first part of the simulation, his explanation changed into a description of heat as molecules vibrating faster. Next, Eric needed to realize how molecular vibrations made the other molecules move. However, here we share what occurred when the interviewer used only planned scaffolding moves.

**Interviewer:** What I’m going to do is pause this for a moment, and I am going to connect these (clicks on connecting bar on the simulation), imagine that this is like, if this was the bowl part of the spoon on the left and if this was the handle part that now we
have more metal of the handle connecting them. So, what do you think you're going to see now when I hit play?

**Eric:** The left are there vibrating, and they also go through the metal until it will go through the metal until it reaches the right.

In this segment, we see that the interviewer followed the prompts from the interview protocol where he established the link between molecular representation and the silver spoon and asked for a prediction. In response, we find that Eric’s prediction is a partial explanation that he maintained after he watched the simulation, and again post-intervention. Even when the interviewer played the simulation for him, he repeated his response, clearly not seeing the molecules influencing the movement of other molecules in the simulation. We think that Eric’s prediction influenced what he noticed in the simulation and thus prevented him from noticing collisions. In terms of scaffolding, the interviewer’s moves directed Eric to use the simulation as a source for confirming his conception. No other prompts were used to direct him to use the simulation as a resource for constructing a new explanation. We think that this case provides a useful baseline for considering the effects of the other scaffolds. For example, viewing these prompts in contrast to the scaffolding Andrew received, we find that none of the five moves used with Andrew were used here. The two moves used here were 1) to establish the link to the spoon and 2) to ask for a prediction. Andrew received these two moves as well, but developed significantly with the other five moves. Thus, we are even more convinced of the need to focus student attention to single molecules moving, and to cue students to think causally.

**Irene – Full explanation, partial and multiple interventions**

We turn to Irene to further investigate these scaffolds. We see that she did not achieve a full score during the simulation, but she did so after more interactions with the interviewer post-simulation. Since these scaffolds were similar in form to Andrew’s questions we share her interview to explore how these scaffolds were enacted in a different way for her. On examining the interaction with the simulation playing, Irene was asked to describe what she noticed. Once again, she did not notice molecular collisions, but she noticed increased wiggling and the changes in the temperature gauges. At this point, the interviewer intervened with a question resembling one of Andrew’s questions.

**Interviewer:** So maybe look at one of the molecules over here (points to the left block of molecules) and try to see, what do you think is making these molecules over here (wiggles the mouse pointer over the right block of molecule) start to move faster?

**Irene:** This side (points to the left block) is like pushing them (shifts index finger sideways a few times towards the right block) to the other side so like this side can start to move (points to the right side).

This is a new conceptual observation for Irene that we think was triggered by the interviewer’s question. In this case we see that the interviewer refocused Irene’s attention towards a single molecule. He then followed up by directing her attention to the causal agency of molecules by asking her “What is making these start to move faster?” We get a more causal-mechanistic response from Irene where she attributes molecular pushing as the cause of more movement. However, unlike Andrew, no further moves were made to reinforce this revelation, and given that her post simulation explanation is only partially complete, this idea was apparently not fully taken up by Irene. However, on looking beyond the simulation, we noticed the interviewer followed up on her explanation by requesting that she gesture about molecules in two ways. In the first way (marked as SHOW:INT in Figure 3a) the interviewer asked her to close her fist to embody a molecule and asked her the causal question regarding temperature, i.e. “Show me how the hot part makes the cold part hot?” Irene responded to this question by describing how dipping her hand in hot water made it hot, and she described increased molecular movement due to heat. Then the interviewer made a second request (marked as SHOW:PART in Figure 3a) and used his fists to represent molecules with one side wiggling faster than the other, and asking her specifically to describe using her hands how the cold molecules moved faster as the hot molecules bumped into them. Irene adopted this gesture and represented two molecules, one hand as a faster wiggling molecule bumping into a stationery colder molecule, while attributing causation to the movement by stating that “the warm molecule goes and hits the cold molecules, and they both move at the same time”.

In summary, four scaffolds occurred in this interaction. 1) refocusing student attention to the movement of a single molecule, 2) causal questioning, 3) request to gesture to explain, and 4) request to represent molecules with hands to explain. Comparing this to Andrew, we see that all four were used in his interaction as well. Thus,
we see an increased value in providing students context-specific gestures (such as a wiggling fist to represent a molecule) and cueing the mechanism through the use of their fists as important embodied resources that can strengthen their understanding of the molecular mechanisms. Next, we consider these moves again with John.

![Figure 3. (a) Irene’s scores in response to interviewer scaffolding. (b) John’s scores in response to interviewer scaffolding.](image)

**John—Partial explanation, partial intervention**

We bring up this final case to share the nuances of scaffolding that appeared to resemble the previous cases but did not lead to similar results. In John’s case, the interviewer maintained focus on molecules but performed these moves before she played the simulation. We share a segment of this interaction where John is asked for a prediction before viewing the simulation.

**Interviewer:** What do you think would happen once I connected a bar of molecules in the middle over here? Between the hot and the cold?

**John:** I think that all of the hot would get stuck on each of the cold eventually heat it up. (indicates the left block moving towards the right block).

**Interviewer:** But you said that heat was also the moving molecules so, are the molecules doing anything to each other?

**John:** No. It's just the heat because when there is no heat it stays still. So, when there is heat added, the molecules move. So once the hot part of the molecules touch the cold one, it will eventually move the molecules.

**Interviewer:** (plays simulation and John watches it intently).

**John:** So, like I said. Um, kind of. Like that side (points to the left block) is moving a lot more. But, they are kind of getting, slower and slower (shifts hand along the connecting bar). But, the heat is also touching this (points to the right block), which is kind of rising there (points to the right thermometer) which is also moving it.

In this interaction, we see evidence of a few familiar moves. First, much like in Eric’s case, the interviewer asked John to make a prediction. He responded by claiming that when the blocks touched, heat would transfer over to the cooler side. Then, the interviewer reminded him of the movement of molecules, a fact they had established at S1, and directed John’s attention to them by asking the causal question “are the molecules doing anything to each other?”. However, John merely claimed that they would wiggle the same way he had observed during S1. Thus, the interviewer used three scaffolds in this segment. 1) To ask for a prediction, 2) to refocus student attention to the movement of molecules, and the 3) to ask a causal question. These scaffolds are the same moves that were used with Andrew and Irene with the difference being that John was not asked to gesture. However, on examining this case, we think that these three scaffolds did not appear to be asked at the right time. In Andrew’s and Irene’s case, these moves were asked with reference to events occurring in the simulation. They were not questions investigating the student’s conceptions. On the other hand, in John’s case, these questions were directed at his current conception which was that heat is a substance that travels through the molecules making them wiggle more. Therefore, much like Eric, John was not directed to use the simulation as a constructive resource for developing an explanation. It seemed to be used as a resource to confirm his current explanation.

**Discussion**
When we review how we facilitated student interactions with a computer simulation, we noted seven unique scaffolds asked across these students. They are 1) establishing the link to the spoon, 2) asking for a prediction, 3) refocusing student attention to the movement of molecules, 4) causal questioning, 5) providing direct instruction about mutually established facts, 6) requesting for gestures to explain, and 7) requesting to represent molecules with hands to explain. Andrew was asked all seven questions, followed by Irene who was asked six. John was asked three of these questions while Eric was asked two. However, we do not think that a higher amount of scaffolding led to sophisticated student explanations. These cases revealed that these moves are time and context dependent, and that changing the order of questions did not lead to better explanations. For example, John was asked causally about molecules, just like Irene. However, since his questions were asked before the simulation was played, they did not have the same effect for him as it did for Irene. Moreover, we found that requests for gestures were productive for Andrew and Irene where not only did they ground the molecular motion in gestures, but they reified the causality of molecular collisions by helping students reflect on their ideas after viewing the simulation. This work shows how scaffolding interactions with science simulations by synergistically using verbal, visual, and gestural resources can lead to sophisticated causal explanations. While successful science environments benefit from synergistic scaffolding (Reiser & Tabak, 2014), creating the right balance requires careful and timely synthesis using multiple representations and modalities. We have shown how causal questions and gestures contribute to this synergy. Future work can examine this work in other scientific practices as well as with new immersive technologies.

References

Acknowledgments
Many thanks to Robert E. Wallon and Robb Lindgren for their conversations throughout this process. This material is based upon work supported by the National Science Foundation under grant no. DUE-1432424. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Scaffolding Authentic Wearable-Based Scientific Inquiry for Early Elementary Learners

Virginia L. Byrne, University of Maryland, vbyrne@umd.edu
Seokbin Kang, University of Maryland, sbkang@umd.edu
Leyla Norooz, University of Maryland, lnorooz@umd.edu
Rafael Velez, University of Maryland, rvelez@umd.edu
Monica Katzen, University of Maryland, mkatzen@umd.edu
Afe Addeh, Eleanor Roosevelt High School, addehafe@gmail.com
Jon Froehlich, University of Maryland, University of Washington, jonf@umd.edu
Tamara Clegg, University of Maryland, tclegg@umd.edu

Abstract: Wearable sensors show promise in engaging youth in scientific inquiry by leveraging physical activity for life-relevant inquiry. Prior research, however, has found that elementary school-age children struggle with the learner-determined aspects of an authentic scientific experiment (e.g., asking a testable research question). We explore how to integrate wearable-based inquiry into early-elementary classrooms with Live Physiological Sensing and Visualization tools, including: designing age-appropriate scaffolds, adapting to teacher perspectives, and meeting Next Generation Science Standards (NGSS). We conducted a two-year, iterative process to develop scaffolds and implement them in four elementary classrooms (two teachers; 90 children). We present two case studies to demonstrate how our participatory designed scaffolds impact the authenticity of the learners’ wearable-based inquiry experience. Our findings contribute insights about facilitating wearable-based inquiry with elementary learners and specific supports for using sensor-based learning systems to meet NGSS.

Introduction

With wearable sensors, such as heart rate monitors and accompanying visualizations, upper elementary- and middle-school learners can design and conduct life-relevant experiments using their bodies (Clegg, et al., 2017; Lee & Drake, 2013; Lee, Drake, & Williamson, 2015; Schaefer, Carter Ching, Breen, & German, 2016). We refer to this emerging area as wearable-based inquiry (WBI). Unlike traditional classroom science experiments, in WBI, learners explore scientific concepts (e.g., exercise, stress) that are life-relevant. In this way, wearables enable learners to scientize their everyday lives (Clegg & Kolodner, 2014)—e.g., their physical activities, body movements, and physiological responses. Grounded in embodied learning theory, which emphasizes the role of the body in facilitating learning (Lee, 2015), our work aligns with the Next Generation Science Standards (NGSS; 2013) and the National Research Council’s (NRC; 2012) efforts to make classroom science more inquiry driven. While existing literature has addressed the use of WBI at the fifth-grade level and higher (e.g., Lee & Thomas, 2011), we could find no prior work on how early-elementary learners (first- through fourth-graders) engage with wearables for scientific-inquiry nor the socio-technical scaffolds necessary to support their learning. Prior work in other contexts (Beyer & Davis, 2008) has shown that early learners require consistent and clear scaffolding and struggle to articulate scientific explanations and predictions, and develop procedures. Our aim is to design and evaluate scaffolds that help early learners acquire the skills necessary for conducting WBI. This aim is partially motivated by NGSS (2013), which states that learners should be able to plan and conduct science investigations before the fifth-grade. Our research addresses the following questions: (1) How do scaffolds impact the authenticity of children’s scientific WBI across grade levels? (2) How can we design multi-dimensional scaffolds for WBI that integrates technology tools, peers, facilitators, and paper-based materials?

To address these questions, we present an evaluation of scaffolds used to support early learners in conducting authentic scientific WBI projects with two wearable sensing tools: BodyVis and SharedPhys (Figure 1). These tools leverage body data for real-time inquiry through Live Physiological Sensing and Visualization (LPSV; Norooz, et al., 2016). Children wear a physiological sensor and draw evidence-based conclusions about their bodies by analyzing visualizations of their live heart-rate data as displayed on an electronic textile shirt (BodyVis, Figure 1a) and a large-screen display (SharedPhys, Figure 1c; Kang, et al., 2016). Our research team conducted two one-week deployments of these WBI tools (four in-class sessions across three classrooms): one in 2016 and another in 2017. Here, learners progressed from semi-structured activities (e.g., brainstorming and testing physical activities to see the impact on heart rate) and examples to planning and conducting their own inquiry projects. On Days 3 and 4, learners asked their own questions such as, “How does my heart rate change when I do the Carlton dance?” or “How does my heart rate change when I do my homework?”
Through a design-based research approach (Sandoval & Bell, 2004), we iteratively designed scaffolds using participatory design with teachers, classroom deployments, and teacher and child feedback. Using a multiple case-study methodology (Yin, 2014), we identify similarities and differences among the authenticity of WBI projects and the influences of our iterated scaffolding from 2016 to 2017. Our findings illustrate how researchers can construct age-appropriate scaffolds for early learners conducting WBI.

Related work
Prior work has shown that early learners can conduct their own meaningful life-relevant science inquiry (Clegg & Kolodner, 2014). Our design is therefore based on Chinn and Malhotra’s (2002) framework for scientific inquiry, which highlights learners’ freedom to select questions of interest. On a national level, elementary-age learners are encouraged to design and conduct inquiry projects that reveal patterns in their environment and make evidence-based claims (NRC, 2012; NGSS, 2013). Using a constructivist approach, we believe scientific inquiry is powerful for young children because it leverages their pre-existing knowledge and enables them to ask questions about their daily lives and bodies that then drive the inquiry process as opposed to directing their inquiries to predetermined investigations (Chinn & Malhotra, 2002; González, Moll, & Amanti, 2005; Hmelo-Silver, Duncan, & Chinn, 2007). Our embodied approach (Lee, 2015) specifically links these aspects of children’s scientific inquiry, as well as their physiological and mathematic learning to their everyday body movement.

Prior work emphasizes that early learners can reach tasks that they could not independently achieve with scaffolding because the work is guided by, for example, paper-based materials (e.g., worksheets) and other knowledgeable people (e.g., facilitator- and peer-based scaffolding; Carter-Ching & Kafai, 2008; Hmelo, Holton, & Kolodner, 2000; Wood, Bruner & Ross, 1976). Facilitator-, peer-, and paper-based scaffolding prompts learners to consider new problem-solving strategies and reminds them of the strategy when they need assistance during the inquiry project (Reiser & Tabak, 2014). Additionally, technology-based scaffolding, as suggested by Quintana et al. (2004), can help scaffold complex concepts to build on the learner’s intuitive understanding. We apply and adapt these scaffolding approaches to WBI, specifically in supporting the scientization of everyday activities with wearable sensors (Clegg & Kolodner, 2014; Metz, 2004; NRC, 2012).

Method
Through a design-based research approach (Sandoval & Bell, 2004), we iteratively designed tools, activities, and scaffolds to support children’s learning about body systems and scientific inquiry through WBI experiences. In both 2016 and 2017, we partnered with three elementary public-school teachers. In each teacher’s class, we conducted four, one-hour sessions of our wearable-based learning program with LPSV tools (BodyVis and SharedPhys; Table 1). To understand the range of elementary-aged children’s experiences, we analyzed cases from two of the three classes—the youngest group (first-grade) and the oldest group (fourth-grade). We present a case study of the 2016 deployment (45 learners) followed by an analysis of and updates to our scaffolds, which we analyzed as an additional case study in the 2017 deployment (45 learners). Both sequences were conducted at the same public elementary school in the Washington, DC metro area (68% African American, 23% Hispanic/Latino, 3% Asian, 2% Caucasian, 4% Mixed Race; 65.6% free and reduced-priced meals).

We designed the initial learning activities, materials, and scaffolds of our 2016 deployment to align with NGSS (2013). Before implementation, we solicited feedback from our teacher partners. After the 2016 deployment, we conducted a participatory design session (Fails, Guha, & Druin, 2013) with the teachers to design the 2017 deployment. The result yielded new learning activities and goals, assessments, and facilitation plans that
enhanced alignment with both NGSS (2013) and NRC inquiry standards (2012). We collected audio, video, and prototyping artifacts from participatory design sessions with the teachers as well as post-session notes.

For the case studies specifically, we analyze changes among the authenticity of the WBI small group projects and the influences of our redesigned scaffolds from 2016 to 2017. In each case, we observed five groups across two classrooms: three first-grade groups and two fourth-grade groups. For each group, we collected video and audio data, photographs, and inquiry project artifacts (e.g., easel pad paper, drawings, worksheets). We also performed post-study interviews with the teachers to better understand their perspectives of WBI tools and learning activities and to solicit design suggestions. All interviews were transcribed.

Our data analysis is based on the NGSS (2013) and Chinn and Malhotra’s (2002) framework for authentic scientific inquiry which breaks the scientific inquiry process into components (e.g., asking questions, developing hypotheses) and describes each in terms of simple-to-authentic scientific inquiry. Using this framework, we designed a codebook to help us identify, for example, how and when students discuss controlling variables and instances in which children discussed their data (NGSS, 2013), as well as the roles of different environmental actors (e.g., facilitators, peers). We then deductively coded video and audio of small groups conducting inquiry projects (Saldaña, 2015). Two researchers independently coded the same videos for one 2016 group, then met to discuss disagreements and clarify code definitions. For the remaining videos, researchers selected and coded videos of a random sampling of two to three groups across classrooms. This process was repeated for 2017. Coders then discussed and summarized major themes.

Teacher interview transcripts and the audio and artifacts of the teacher design session were structurally coded by topic (e.g., design suggestion for learning activities, classroom ecosystem; Saldaña, 2015) by two researchers. After discussion, the researchers did a second round of coding to identify themes (Saldaña, 2015), and then synthesized the themed data for the larger research team to review. We compiled our summaries and findings into two cases (one for 2016, one for 2017) and performed axial coding to identify themes across groups by year (Saldaña, 2015). By grouping themes by year, we could look across the two iterations to identify differences in how learners designed and conducted their WBI projects. Our cross-case analysis thus considers the ways our iterated scaffolds (Table 2) influenced the learners’ authentic inquiry experiences.

Findings
We present six themes regarding struggles learners faced in 2016 that we addressed with scaffolding changes in 2017. First, a vignette from our 2016 case illustrates the identified challenges. Then, we provide a description of the themes and map the scaffolding changes to ways in which learners’ WBI projects changed in 2017 with respect to scientific authenticity. We demonstrate the influence of these scaffolds by presenting two 2017 vignettes, which illustrate the impact of the scaffolding across grades and spread among facilitators, peers, and materials.

2016 deployment findings
In 2016, first- and fourth-graders were engaged with their inquiry projects and could follow the facilitator’s directions, however, they needed scaffolding to conduct WBI projects and not merely play with the LPSV tools.

On Days 3 and 4, first-grade Group B’s assigned facilitator explained how to conduct a scientific experiment, facilitated their brainstorming and decision making, and assisted them in recording their ideas. Sitting in a circle, the facilitator began the experience by asking each learner to share a question they wanted to test with the LPSV tools. Learners struggled to either think of or share ideas on the spot. They instead silently reached for markers to draw on chart paper. When they did share, learners discussed how the physical activity would be done instead of the impact on their heart rate (e.g., addressing who could run the fastest instead of the impact of running on heart rate). Often these ideas could not be tested with the LPSV tools or within the classroom. The burden was on the facilitator to determine what was testable, but she did not share her criteria. For example, when a child asked to measure the heart rate of their pregnant teacher, the facilitator said no but did not explain her rationale. Instead, she gave suggestions for questions, one of which the group selected to test (e.g., the impact on heart rate

<table>
<thead>
<tr>
<th>Day</th>
<th>Session Activities for both 2016 and 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Learners used BodyVis to identify organs in the body and to conduct group inquiry with the heart and breathing rate functions. Facilitators solicited learners’ questions about anatomy and physiology.</td>
</tr>
<tr>
<td>2</td>
<td>Facilitators led semi-structured activities in which learners used SharedPhys to conduct experiments.</td>
</tr>
<tr>
<td>3</td>
<td>Using their testable question from the previous day, learners designed procedures, crafted a hypothesis, and ran a trial of their procedures. Each group recorded their experiment on a sheet of chart paper.</td>
</tr>
<tr>
<td>4</td>
<td>Learners conducted their experiment, interpreted results, and presented findings to the class.</td>
</tr>
</tbody>
</table>

### Table 1: High-Level Description of Learning Activities for First- and Fourth-Grade Classrooms Deployment

<table>
<thead>
<tr>
<th>Deployment</th>
<th>Classrooms</th>
<th>Grade</th>
<th>Deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>66</td>
<td>65</td>
<td>64</td>
</tr>
<tr>
<td>2017</td>
<td>65</td>
<td>66</td>
<td>67</td>
</tr>
</tbody>
</table>
when running across the room). When the group conducted their experiment, they became so excited about which child would “win” (i.e., have the highest heart rate) as displayed by verbal and physical cheering, they did not discuss the overall data trend.

Across groups, we observed limitations in the authenticity of learners’ experiments that were exemplified by this group. First, learners struggled to articulate their ideas into a testable question without a facilitator to rephrase their ideas. The facilitator also did not reveal a rationale for guiding the group towards a particular question. Thus, while learners were excited by their ideas (e.g., who could run the fastest), they overlooked how the LPSV tools could be used in experimentation. We also observed that children struggled to interpret SharedPhys’s visual line graph to make claims. During the SharedPhys experiments, children often became so focused on those performing activities (e.g., running) that they did not pay close attention to the live data visualization—which only showed the last one-minute of data and could not be paused. This became a challenge when attempting to draw conclusions because children could not actively refer to the graph during discussions.

After the experiments, students did not make theoretical claims, rather they observed without asking why or discussing the connection between physiology and their WBI projects. The burden was on facilitators to connect the experimental results to the human body. All scaffolding was provided by just-in-time facilitation in the form of question prompts, reminders about next-steps, and feedback about learners’ ideas for experimental questions or procedures. We recognize that this high-touch facilitation model would likely be intractable in a typical classroom where student-to-teacher ratios are much higher. Though the above vignette focuses on first-grade, we identified similar issues in the fourth-grade class; however, we found that fourth-graders were more comfortable in developing hypotheses and, with facilitator support, interpreting line graphs.

**Iterating scaffolds for 2017 deployment**

Given our 2016 deployment findings, we iterated our scaffolds to support more authentic WBI experiences and to reduce adult facilitation requirements. First, we reviewed the NGSS on how elementary learners should be conducting experiments. We then partnered with our teachers to design facilitation prompts and materials. Our analysis of teacher interviews and design ideas identified six key themes:

**Constrain research questions to a set of testable criteria.** In 2016, it was apparent that learners misunderstood the constraints of the LPSV tools by suggesting questions that moved outside of the classroom, were untestable within the time limit (e.g., sleeping), or otherwise problematic (e.g., taking poison). In 2017, we presented learners with a visual guide (Figure 2a) and testable question criteria (Table 2). Learners were then asked to self-check their questions with these criteria.

**Vocabulary definitions (e.g., increase, decrease) for language to express ideas in a testable frame.** In 2016, we observed that learners struggled to put words to their science questions and hypotheses. In 2017, we introduced an active learning vocabulary lesson and posted the words on the wall for the duration of our deployment (Figure 2b). In addition, we provided the first-graders with worksheets containing a word bank and

<table>
<thead>
<tr>
<th>Scientific Inquiry Stage</th>
<th>2016 Scaffolds</th>
<th>2017 Scaffolds with Developmentally-Appropriate Scaffolds by Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Development</td>
<td>Anatomy vocabulary lesson consisting of a large group discussion and scavenger hunt.</td>
<td>Both classes were provided with an anatomy and inquiry-related vocabulary lesson consisting of a class discussion, scavenger hunt, slideshow, and posters of words. Criteria were set for creating a testable question—the procedure for testing the question needed to be: conducted within one minute, conducted in the classroom with available materials, and measurable with an LPSV tool. First-graders were also provided with embodied definitions, a word bank on all worksheets and an assigned a facilitator who used sentence framing cards to aid learners to rephrase their ideas. Fourth-graders collaborated through online writing activities.</td>
</tr>
<tr>
<td></td>
<td>Assigned facilitators assisted brainstorming and rephrased ideas into a testable format.</td>
<td></td>
</tr>
<tr>
<td>Procedure Design</td>
<td>Assigned facilitators led discussion and poster recording.</td>
<td>Both classes were provided with a chart paper with empty boxes with headings for each stage of the inquiry process. First-graders used a graphic organizer worksheet to draw and annotate ideas before facilitated discussion. Fourth-graders had access to floating facilitators for feedback.</td>
</tr>
<tr>
<td>Data Interpretation</td>
<td>Conducted experiments as a class on Day 2.</td>
<td>During Day 2 large group experiments, facilitators taught embodied methods of analyzing changes in SharedPhys’ visualization. First-graders practiced interpreting line graphs by drawing their hypotheses for the worked example and comparing it to the results.</td>
</tr>
<tr>
<td>Theory Building</td>
<td>Facilitators led large-group discussions.</td>
<td>Facilitators synthesized findings after experiments were conducted and led a large group discussion. First-graders were provided with a poster on which the facilitators recorded the results of all experimental findings.</td>
</tr>
</tbody>
</table>
sentence structure cue cards. We also taught first-graders embodied definitions (i.e., Total Physical Response; Asher, 1969) for words beyond the NGSS (e.g., circulatory system). These body movements ground complex physiology content in gestures and meanings familiar to children (Lee, 2015).

Figure 2. WBI Scaffolds: (a) testable question criteria, (b) vocabulary posters, (c) paper-based drawing worksheets, (d) experimental findings posters, and (e) large chart paper with a WBI template.

Grade-specific writing activities to aid idea expression and procedural thinking. Noting how learners of both grades struggled to develop questions and experimental procedures, we designed writing activities to reduce the burden on the facilitators. As recommended by the teachers, we took a grade-specific approach. We provided first-graders with a three-panel graphic organizer worksheet on which they could draw and write before being asked to verbalize their ideas (Figure 3). Graphic organizers guided learners to make chronological sense of the procedures by having them describe steps in order. By using drawing instead of text to explain abstract ideas, we bridge the learner’s intuitive understanding of complex concepts to the inquiry process (Quintana et al., 2004). For the fourth-graders, learners collaboratively brainstormed research questions, shared knowledge, and remixed each other’s ideas through online collaborative writing activities (in Google Docs). The fourth-graders discussed their ideas and received peer feedback.

Opportunities to practice prediction and interpreting line graphs. In 2016, we found that while fourth-graders could develop hypotheses and interpret line graphs, first-graders struggled to connect the SharedPhys visualization to their abstract understanding of physiology. In 2017, we scaffolded hypothesis development by including discussions, embodied interpretation (e.g., using their arms to trace the line in the air; Asher, 1969), drawing, and recording after each experiment. Realizing their need to doodle or draw their understanding before discussing it, we asked first-grade children to imagine performing an activity and then draw a prediction of their heart rate on graph paper resembling the SharedPhys screen (Figure 2c). We facilitated discussions on learners’ graphs and corrected misinterpretations (e.g., Maya is the blue line, Ed is the green line). We encouraged children to use embodied definitions of increase and decrease by physically acting out their predictions (Asher, 1969). Using paper arrows, we posted the learners’ predictions onto the SharedPhys screen before learners wearing the sensors (i.e., wearers) did the activity. During the experiment, the class could observe how the real-time heart rates change compared to their prediction heart rates. Finally, we asked wearers to describe what their heart rate felt like. Wearer testimonials and SharedPhys results were then compared to the predictions.

Synthesis of results across groups for collective understanding and building theory. To bridge connections across experiments and reaffirm understanding, facilitators rephrased the findings of each experiment after the class had drawn their own conclusions and supported learners in theory building in a way that was developmentally-appropriate. In 2016, first-graders often could not state the findings of a past experiment. So, in 2017, facilitators recorded each first-grade group’s experimental findings on a large poster so that learners could see and reference prior findings (Figure 2d). Fourth-graders were better able to remember prior findings in discussions but facilitators helped them to synthesize data from across the experiments into a theory.
Visualization of the inquiry process so that learners are aware of next steps. In both years and in both classes, small groups were given markers and large chart paper (Figure 2e) to record their names, research question, hypotheses, procedures, heart rate data, and findings. In 2017, however, facilitators drew boxes on the chart paper corresponding to the stages of WBI for learners to fill-in. This template was meant to guide learners through the WBI process by revealing what they needed to complete and record at each stage.

2017 deployment findings
In 2017, we observed that with our iterated scaffolds, early elementary learners needed less high-touch facilitation to develop a testable question and procedures, and accurately interpret visualized data into evidence-based theory. They still needed more facilitator guidance, however, when synthesizing results across experiments. To demonstrate the impact of our scaffolds, we present two vignettes:

Vignette 1: First-graders investigating the effect of the Nae-Nae dance on the body
After providing learners with testable question criteria, they brainstormed activities that were personally meaningful and met the provided criteria. One boy, Alex, shared his excitement to test dancing because of his love of dancing—especially doing the Nae-Nae. The facilitator, Omar, asked the group for feedback and they agreed to investigate the impact of doing the Nae-Nae dance on their heart rate.

The three-panel drawing worksheet allowed learners to externalize abstract ideas into a sequence of drawings and annotations (Figure 3). Without facilitation, many students used the three panels to draw a sequence of observation – action – observation. Omar presented learners with sentence structure cue cards with ideation and sentence framing words (e.g., If, Then) to enable them to format their ideas into a procedure for investigating causal relationships. These cards did not appear to be impactful, however, because the learners paid little attention to the cards and continued to draw their ideas. After drawing and writing on their individual sheets, the children shared their ideas, formalized their procedure, and wrote it in the inquiry template on the large chart paper.

As a class, each group conducted their experiment and Omar used the poster (Figure 2d) to visually display each group’s findings. We observed that learners referred more to past findings to make claims than in prior years, often while looking at or pointing to the visual repository. This visual repository made it easier for learners to refer to past findings to make claims and for Omar to bridge together findings to pose theories. After the Nae-Nae experiment, Omar wrote “increased a little” because the class concluded that the dancing increased the wearers’ heart rate but not to the same extent as some prior activities. After all experiments were done, Omar had the class refer to their poster of results to synthesize their findings into a theory (e.g., “activities like running and galloping lead to a bigger increase in heart rate”). When asked what they learned, children volunteered new theories (e.g., when you exercise big muscles like the legs, they need more oxygen and so your heart beats faster), which helped demonstrate their grasp of the relationship between physiology and activity.

Vignette 2: Fourth-graders investigating the effect of being scared on the body
In the fourth-grade class, facilitators floated between groups instead of being assigned to a group as in the first-grade class. Group A needed facilitator attention because they could not agree on a physical activity to test. The facilitator, LaSonya, in trying to mediate the group tension, talked about how emotional reactions (e.g., laughing, feeling scared) also impact our heart rate. After some debate, the group decided to test how being scared affects heart rate. Group A’s conversations focused on completing each section of the chart paper WBI template and imagining how much their idea would impress their peers. Without prompting, Group A identified ideas that were not testable (e.g., “Seeing a scary movie would take too much time”). Group A’s ideas for how to scare someone were shared as personal stories from when they remembered being scared (e.g., “We should do the thing my sister did...”). They generalized from life experiences when they were most scared and developed a list of ways to scare the wearers. The group was able, then, to narrow this list based on what fit the testable criteria.

For their SharedPhys experiment, the group tested three scare tactics (e.g., being poked, grabbed, and screamed at) and observed the wearer, Janet’s, resulting heart rate. In response, however, Group A’s classmates had concerns about the validity of the experiment. They questioned if the increase in heart rate was caused by Janet jumping or laughing instead of being scared (e.g., “She was laughing which got her heart rate up”). Group A was also skeptical of the accuracy of the SharedPhys data and voiced concern that they could not identify on the graph when each scare tactic was implemented.

2017 case summary
These vignettes illustrate how learners asked life-relevant, testable questions and developed their own experimental procedures. Our findings suggest that providing testable question constraints lessened the complexity of asking inquiry questions and planning procedures for learners (Quintana, et al., 2004) while also
offering freedom to explore life-relevant investigations. This was evident in our observations of learners using the testable criteria to discuss and decide among themselves which ideas and procedures were testable. Our analysis of first-graders’ discussion contributions and drawings suggests that learners in our 2017 case had a clearer understanding of how to interpret the visual line graphs in SharedPhys. We attribute their increased interpretation of line graphs to the drawing scaffolds in which learners mapped their own visual hypotheses to the real-time results visualized on SharedPhys (Figure 2c). Lastly, after conducting the embodied experiments, first-grade students made theoretical claims about the connection between physiology and physical activities—a key component of the Chinn and Malhotra (2002) framework that we did not see in the 2016 case. We did not observe a difference in how fourth-graders across both years developed theories across experiences, suggesting that the LPSV tools and 2016 activities were enough to help them make connections to physiological phenomena. While facilitators still had to help learners develop questions and procedures in their small groups in 2017, they were able to make use of paper-based scaffolds (e.g., chart paper, worksheets) that reduced the burden on how much help they needed to provide to each group. Facilitators provided more scaffolding in 2017 during whole group conversations for recording observations and synthesizing results across experiments.

Discussion

Although some prior work in WBI notes the importance of scaffolding children’s use of technology for inquiry (e.g., Lee, Drake & Williamson, 2015), we provide the first examination of the impact of scaffolds on the inquiry process. Through our iterative design process, we considered how multiple dimensions of scaffolds (i.e., tools, peers, facilitators, and paper-based materials) could be integrated to support early-elementary learners’ WBI. By integrating facilitation-based scaffolds, paper-based scaffolds (e.g., the prediction drawing worksheet), peer-based scaffolding, and the wearable tools themselves, we found that learners could test and accurately interpret investigations of two types of questions. First, changes in heart rate over time (e.g., change in heart rate when doing the Nae-Nae dance for a minute) and, second, changes in heart rate across learners or activities (e.g., change in heart rate for three different scaring tactics). Paper-based scaffolds bridged facilitator- and peer-based scaffolding so that after some guidance, peers could support each other, compare drawings, and self-check the feasibility of an idea. When learners supported each other in groups, facilitators could focus on synthesizing past findings on large sheets of paper, which learners referred to when trying to generalize across experiments. This theory building process is a key component of Chinn and Malhotra’s framework (2002) and the NGSS (2013).

With respect to authentic inquiry, we found that, as in prior work with older children (e.g., Lee & Drake, 2011; Lee & Thomas, 2011), early elementary learners could design and conduct life-relevant WBI by examining everyday activities (e.g., dancing, scaring, laughing). Although many of the WBI projects in our study included gym or recess activities like prior embodied WBI research (e.g., Lee et al., 2015), three of the five 2017 small groups we analyzed also included other activities such as homework and dancing. We found that, with our integrated scaffolds and LPSV tools, early learners explored the embodied connection between physical activity data and the physiological phenomenon of their bodies (e.g., the circulatory system). Beyond the scope of prior WBI studies, which emphasized interpreting trends in visualized data (e.g., Lee & Thomas, 2011), learners in our study recognized the physiological impacts of stress and fear. Our multi-dimensional scaffolds supported learners’ exploration of their physiological functioning and ability to relate findings back to their own everyday activities.

We also observed limitations in our current scaffolds. We found that despite our integration of differing scaffolds, first-graders still needed facilitator help with literacy-based aspects of the inquiry process (e.g., articulating questions and evidence-based claims). Future research is needed to further enhance such scaffolds for WBI. Additionally, as is common among design-based research studies with multiple sources of data, we had to exclude some details of our study that would have been included in a longer article. Finally, our vignettes represent some of the more life-relevant WBI projects. Half of the groups in each class tested comparisons of physical activities (e.g., galloping, lunges). We therefore need additional scaffolds to help learners more systematically develop creative questions unique to their specific interests and experiences.

Our findings contribute to how educators can implement wearable sensors for inquiry and embodied learning in early elementary classrooms and more broadly, how educators can use sensor-based learning to meet NGSS (2013). Our work differs from and expands on prior WBI work by focusing on early-elementary learners, iterating and comparing inquiry scaffolds across two deployments, and working with teachers to explicitly align with NGSS. While we found that a multi-dimensional scaffolding approach can support early learners in planning and conducting authentic WBI, more work is needed to understand how peer scaffolding can be leveraged to lessen the burden on educators so that WBI is more realistic to the typical student-teacher ratio.

References


Acknowledgments
This work was supported by NSF Grant (IIS-1441184). We thank the participating children and teachers.
How Teachers Use *Instructional Improvisation* to Organize Science Discourse and Learning in a Mixed Reality Environment

Maggie Dahn, University of California, Los Angeles, maggiedahn@gmail.com
Noel Enyedy, University of California, Los Angeles, enyedy@gseis.ucla.edu
Joshua Danish, Indiana University, jdanish@indiana.edu

**Abstract:** In this paper we describe our model of *instructional improvisation*, which blends rules of theatrical improv with a constructivist teaching approach to suggest practical moves teachers can make during classroom lessons to support students’ science inquiry. We describe instructional improvisation through an analysis of interaction in the STEP (Science through Technology Enhanced Play) environment, a mixed reality simulation in which young children (ages 6-8) learn about complex science concepts. We present a case study analysis (n=26) and demonstrate how teacher moves aligned with our model of instructional improvisation support joint ownership of science knowledge and students having agency within science learning. After demonstrating the effects of these moves, we end with a discussion of how teachers might intentionally use instructional improvisation to support productive interactions in inquiry-based science classrooms.

**Introduction**

Good teachers develop plans to help students learn course content, yet constructivist theory requires they also recognize the value of students’ prior knowledge and encourage students to be active participants in their own learning (Sanchez & Valcárcel, 1999). Thus, when students’ interests and previous experiences lead to unexpected moments in the classroom, teachers must deviate from their scripted plans. While it is true that students can actively construct knowledge from any experience, including lectures and videos, it is hard to predict in advance just how instruction will interact with a child’s existing knowledge. Seemingly clear statements can be assimilated into a child’s existing understandings in unpredictable ways that distort the intended meaning to fit with what they already understand. Researchers have used improvisation as a metaphor to describe how teachers strike a balance between creating structures and being responsive to students during moments of instruction (Erickson, 1982; Sawyer, 2011).

Improvisation of this form is not simple nor is it entirely spontaneous. Audiences of theatrical improv and popular television shows such as *Whose Line Is It Anyway?* can be led to believe that improvisation is something that just happens between actors without much prior planning or thought (Yanow, 2001), however, theatre professionals describe improvisation as an overarching framework with specific rules that guide interaction (Fey, 2011; Halpern, Close & Johnson 1994; Sawyer, 2004). Prior work on improvisation and teaching has used existing rules of theatrical improvisation as a framework for identifying aspects of effective teaching and curriculum planning (Borko & Livingston, 1989; Brown & Edelson, 2003) or as a set of theatrical activities for teachers to engage with and apply to their practice (Lobman & Lundquist, 2007). However, the rules of improv do not neatly map onto the types of interactions that lead to student learning because teachers in classrooms have different goals than actors in improv theatre. Sawyer (2004) draws from improv to suggest practical ways for constructivist-minded teachers to integrate improv within their teaching. We aim for a more precise articulation of this translation and propose a hybrid model—*instructional improvisation*—of teaching as improv that aims to explicitly blend the rules of theatrical improv with research on constructivist teaching within science classrooms to suggest moves teachers can make to support students’ science inquiry. Our model can help teachers remain flexible by offering guidance in how to respond to student contributions throughout a lesson and engage the ensemble with emergent concepts. We describe the model of instructional improvisation through a case study of one classroom (n=26) in our context: the Science through Technology Enhanced Play (STEP) learning environment (Danish et al., 2015; Enyedy, Danish, & DeLiema, 2015), a mixed reality simulation used to teach children about science concepts.

**Rules of improvisational theatre in an education context**

The rules of theatrical improvisation are established, yet fluid. For our model we developed and adapted a composite of six rules drawing from advice of improv experts (Fey, 2011; Halpern, Close & Johnson 1994; Sawyer, 2004). The rules of our model of instructional improvisation include: 1. Always agree; 2. Yes, and…; 3. Make statements or ask questions that elicit statements; 4. No mistakes, only opportunities; 5. The needs of the
ensemble are greater than the individual; and the rule towards which all the others aim, 6. Tell a cohesive instructional story. Below we describe the rules in theatre, focusing on how actors make and receive offers—dialogue or contributions that advance the narrative scene (Halpern, Close & Johnson 1994)—and translate the rules to make the model useful for teachers.

Always agree
The always agree rule is usually introduced as the first rule of theatrical improv because it frames how actors should interact and support one another when developing a scene. To illustrate, if your partner says, “I can’t believe we’re stuck in this dungeon,” you should accept the premise that you are in a dungeon together whether you want to or not. If you instead say, “What are you talking about? We’re not in a dungeon, we’re on a boat,” you have undermined the contribution, making it difficult for your partner to know what to say or do next.

In an education context this does not mean the teacher should agree with every student idea even if it is inaccurate, but it does mean that student contributions should be positioned as legitimate within the collective story the class is trying to tell. If in the case of discussing how bees pollinate a student suggests that bees try to pollinate when they collect nectar, a teacher could “agree” by responding, “That’s interesting. You notice that bees pollinate and collect nectar at the same time.” While the teacher is not agreeing with inaccurate information (bees incidentally pollinate flowers, they do not try to pollinate), she does not evaluate the student’s offer either. Instead, she effectively “agrees” by positioning the student’s everyday sensemaking as a legitimate resource (Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2000) in service of moving the instructional story forward. If she had responded, “I don’t think that’s quite right” or “Are you sure?” she would be undermining the idea, making it difficult for the student to know how to best respond. Because improv is a co-construction of ideas and events, it is important for actors (and in classrooms, teachers and students) to support one another as the scene (lesson) unfolds.

Yes, and...
The next rule of improv is intimately connected to always agree. In addition to agreeing with a partner’s contributions, an actor should also add something, or yes, and... her partner, to move the scene in a forward trajectory. If you only agree that you are in a dungeon but do not add detail—about your relationship with your partner, the plot, or what the dungeon looks/smells like—then you are essentially leaving all of the responsibility on your fellow dungeon dweller to figure out what comes next. A yes, and... move could sound something like, “Yes, thank goodness I brought this spoon to dig us out of here.” In this case, the actor’s offer adds to the narrative the partners are co-constructing as it pushes the scene forward.

In the classroom example above, if the teacher agrees but does not add detail or elaborate further then she would be leaving all of the responsibility on the student to figure out what comes next in their collective inquiry. A yes, and... move could sound something like, “After the bee is pollinating and collecting nectar, I notice that she goes back to the hive. What happens when she gets there?” By using a yes, and... move here the teacher prompts students to add to the story by describing the bees’ subsequent actions. The yes, and... move advances the instructional narrative by linking student contributions to the teacher’s broader learning goals.

Make statements or ask questions that elicit statements
A third rule of theatrical improv is to make statements, meaning that statements usually trump questions because when you ask questions you put pressure on your partner to come up with answers. If in the example above the responding actor asks about the spoon, “How are we going to dig with that thing?” he would be relying on his scene partner to come up with a quick, clever response. Instead, he could say, “Great work. And I have this shovel.” This statement moves the scene forward as the two actors begin to formulate their escape plan.

The make statements rule translates to the classroom a bit differently because it is necessary that the teacher and students ask questions. The main thrust of this rule in our model is that the statements the teacher makes should add to the story and her questions should elicit statements from students that advance the narrative. Regardless of whether the turn at talk is a statement or question, it is important that offers push the lesson forward.

No mistakes, only opportunities
This next improv rule reinforces the first rule of agreement. It requires that as long as the general rules of improv are upheld, there are no mistakes, only opportunities for new discoveries. Because there are no wrong directions, all participants can help the scene evolve. In the classroom, this means that even if students veer down a path to explore a conceptually inaccurate idea, the teacher should allow their interactions to organically unfold because she knows that she can eventually pivot towards the target science content. In fact, she created structures and plans...
before implementing the lesson for exactly this reason. The *no mistakes, only opportunities* rule plays out similarly in both the theatre and classroom because it is all about saying *yes* to whatever curveballs are thrown your way.

**The needs of the ensemble are greater than the individual**

It is important that actors refrain from “stealing” an improv scene in order to have their ideas and jokes heard. Actors should be good listeners and focus on what they can contribute to the narrative rather than writing the script in their heads. Improv is not really about individuals; a successful improv scene is evidenced through the form of the actors’ collective storytelling. This rule of putting the *needs of the ensemble over the individual* is relevant in the classroom because teachers deal with conflicting demands from individual students and must decide which ideas to pursue while simultaneously honoring all student contributions and connecting ideas to core disciplinary concepts.

**Tell a cohesive instructional story**

We end with the most central rule of improv for our model of instructional improvisation, which is to *tell a cohesive instructional story*. In both theatrical improv and improv as used in our education context, the ultimate aim is for participants to tell a story through dialogue that develops relationships between characters and drives the narrative towards a set of co-constructed goals. While not evidenced through isolated moves, the story that the teacher and students tell together is a collective co-construction of knowledge and understanding over the course of a lesson, instructional unit, or other bounded learning experience. Stories make content more memorable (Berk & Trieber, 2009), which is valuable both in the theatre and the classroom. This final rule is one on which all the others depend because all improv moves are made in service of telling a cohesive and compelling story.

---

**Describing the STEP environment**

The Science through Technology Enhanced Play (STEP) environment (see Figure 1) is a mixed reality simulation built to support the collective inquiry of teachers and students in science (Danish et al., 2015; Enyedy, Danish, & DeLimea, 2015). The mixed reality helps participants build conceptual understandings as they link movement in the real world of the play space to what happens visually within the simulated system (Danish et al., 2015; Lindgren, Tscholl, Wang, & Johnson, 2016). In STEP, students participate by stepping into the space and “becoming” bees that retrieve nectar and pollinate flowers on the projected screen. As students move, their motion is tracked by Xbox Kinect cameras using the OpenPTrack software (Munaro, Horn, Illum, Burke, & Rusu, 2014) to feed their location and motion into a computer simulation. We encouraged teachers to be playful in their pedagogical approach but did not explicitly suggest they engage in what we came to understand as instructional improvisation. Our model emerged from observing teachers during STEP lessons given our guiding suggestions to support play-based inquiry.

**Methods**

We designed lessons to teach first and second-grade students about key science concepts, including the differences between nectar and pollen, how bees communicate within the hive to forage for nectar, and the complex process of pollination. While young students might know that bees live in a hive and have something to do with honey, nectar, and pollen, they usually do not know much about how bee communication via a “waggle dance” allows bees to efficiently gather nectar, which in turn supports the larger ecosystem by leading to pollination of the flowers they visit. To illustrate how our model can elucidate teachers’ interactional practices and support play as a form of inquiry, we present a case study analysis of when students first learned about the role of the “waggle dance.” As we watched the teacher allow students to be productively playful, our model of instructional
improvisation emerged as we found ways to name interesting moments in interaction that seemed to support students’ agency in inquiry.

Participants
The STEP project was developed as part of a collaboration across two universities and affiliated elementary schools. The present analysis focuses on the implementation of the bees unit at one school, which involved children from three mixed-age classrooms (n=76; 42 first-graders and 34 second-graders). The three teachers each had more than six years teaching experience and two years working with the STEP system. When we first introduced them to STEP, we emphasized the value of pretend play and allowing students to discover content with minimal guidance. We focused on Ms. Jones’ classroom (n=26) for our initial theory-build because we felt she was particularly successful at balancing students’ agency with the curricular goals in a manner that felt truly “playful” for all.

Data sources and analysis approach
Our goals were to understand how teachers supported the playful, conceptually rich narrative we witnessed and to identify specifically what teachers did to support that narrative. Understanding the teacher’s moves as improvisation emerged during initial video analysis; we then used this frame to analyze the rest of the video data and found similar moves throughout. To determine whether students in the classroom had learned the content, we used pre/post interviews that consisted of ten questions about the target science. Two coders analyzed responses (interrater agreement=97.5%). We developed our model by focusing on Ms. Jones’ class (12 hours). We made logs that indicated where we felt like the teacher was using improvisation, developed conjectures around these points, and adjusted conjectures as we reviewed additional video (Erickson, 2006). We chose two lessons to pursue with the goal of showing how teacher moves encouraged student agency and supported engagement with content.

Findings

Episode 1. Using improv to introduce science content and support student participation
Our analysis of pre/post gains showed there was a significant increase in content learning across classes from pre-test (M=3.8, SD=2.02) to post-test interviews (M=9.18, SD=3.03); (MD=5.38; t(75)=15.47, p < .05), yet the primary focus of our findings is on the model of instructional improvisation in classroom interaction. In the first lesson, Ms. Jones’ objective was for students to understand that bees have the goal of collecting nectar and as they do, they incidentally get pollen on their hind legs, which they bring with them to other flowers. The technology indicated the bees were collecting nectar when they hovered over flowers and a number of hearts rose from the center of the flower (1 heart=okay nectar, 2 hearts=yummy nectar, and 3 hearts=outrageously yummy nectar). As they foraged, the pollen animation was simultaneously activated, meaning that golden sparkles emanated from the flower as the bees collected the nectar. Although students playing bees were not purposefully trying to pollinate, it was impossible to collect nectar without also getting pollen on their legs. The scene below (see Table 1) is from the end of the lesson in which five students played in the STEP space as others participated from the yoga mat hive.

Table 1: Episode 1

<table>
<thead>
<tr>
<th>Turn</th>
<th>Speaker</th>
<th>Talk</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ms. Jones</td>
<td>Adam got some. Dylan’s going …</td>
<td>Dylan walks back to hive</td>
</tr>
<tr>
<td>2</td>
<td>David</td>
<td>You have to fill it and then bring it back to the hive</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ms. Jones</td>
<td>What are you filling it with?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Many students</td>
<td>Nectar!</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Ms. Jones</td>
<td>The honey?</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Jesse</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Ms. Jones</td>
<td>No honey?</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Jesse</td>
<td>Oh, maybe that’s the pollination you did! The-</td>
<td>Leans in from yoga mat</td>
</tr>
<tr>
<td>9</td>
<td>Zed</td>
<td>Oh, pollen!</td>
<td>Points to screen from yoga mat</td>
</tr>
<tr>
<td>10</td>
<td>Jade</td>
<td>Oh, I thought of something. If you like go into there and fill up a lot of nectar</td>
<td>Points to screen from yoga mat</td>
</tr>
<tr>
<td>11</td>
<td>Jesse</td>
<td>When the little dots are coming out of you, that means you’re pollinating</td>
<td>Gestures towards the screen from mat</td>
</tr>
<tr>
<td>12</td>
<td>Ms. Jones</td>
<td>Oooh</td>
<td></td>
</tr>
</tbody>
</table>
This episode begins in line 1 when Ms. Jones agrees with student actions by making statements about what students do with their bodies—collect nectar in the simulation. When she narrates, “Adam got some. Dylan’s going…” she is noting that a key aspect of their behavior is tied to the nectar they collected. By labeling the students’ actions in this way, Ms. Jones focuses students’ attention on how the story is one of collecting nectar and delivering it to the hive, thus tying students’ actions to the building narrative aligned with curricular objectives. Ms. Jones’ labeling is an important first step that supports her next moves.

In turns 5 and 7 Ms. Jones asks students questions that elicit statements—“What are you filling it with?” and “The honey?” Student responses are in the form of three distinct contributions connected to lesson objectives. In turn 8 Jesse predicts that what they see is pollen, not nectar, and Zed verbalizes agreement. In turn 10 Jade connects the location on the flower to nectar collection, and in turn 11 Jesse calls attention to what the animation is communicating—“when the little dots come out of you, that means you’re pollinating.” As students discuss what they know about pollen and nectar, the most important instructional effect of Ms. Jones’ moves is that students build on their prior knowledge and are ultimately able to make distinctions between honey, nectar, and pollen on their own. Of course this episode is just a snapshot of much longer, more involved conversations and debates students had during the first several lessons of the unit. However, Zed’s understanding of the difference is finally made evident in turns 15-17 when he excitedly declares that he “gets it” and provides an explanation that correlates the hearts he sees on the screen with nectar, an understanding that aligns with the “correct science.” In turn 19 other students talk over one another in agreement. Through this short exchange, Ms. Jones’ intentional moves allow students to have the floor to author the class story and participate in the shared ownership of science knowledge.

**Episode 2. Improvisation to support joint construction of science learning**

In the fourth lesson students investigated how the bees’ “waggle dance” works to communicate nectar locations and then invented their own communication system in order to tell one another where to forage for the best nectar. A small group of students would fly into the mixed reality field as bees to find previously hidden flowers and collect nectar. The students would then fly back to the hive (the literal inside of the hive was projected on the screen) and dance for a partner bee to show them where to find a flower with nectar. Like bees, students invented ways to communicate three critical pieces of information about the flower—its distance from the hive, its direction, and the quality of the nectar. For bees, the distance is communicated by the length of the dance itself, direction is communicated in relation to the angle the flower is from the sun, and quality is communicated by the amount of times the bee repeats the dance and how rapidly the bee waggles (our students called it the bee’s “booty shake”). Bees communicate this information to the hive at large when returning from a foraging trip. In the scene below (see Table 2), students in the hive just watched their partner bees’ waggle dances and went to find the flowers. However, they were unable to locate them in the field because instead of translating the dance from inside the hive to out in the field, students went to the literal spot on the floor where their partner ended the dance (see Figure 2).

**Table 2: Episode 2, scene 1**

<table>
<thead>
<tr>
<th>Turn</th>
<th>Speaker</th>
<th>Talk</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Guys, not near the sun!</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Jesse</td>
<td>Go to the sun!</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Dylan</td>
<td>Go to the red flower, David!</td>
<td>A predator flies across the screen</td>
</tr>
<tr>
<td>4</td>
<td>Jesse</td>
<td>David!</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Zed</td>
<td>Ah, I died!</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6 Zed I died! Did you see that?
7 Ms. Jones What happened?
8 Zed The, the eagle came, and I died!
9 Ms. Jones Oh, lay down
10 Ms. Jones Lay down
11 Zed Oh
12 David I can’t find it
13 Ms. Jones Zed’s dead
14 Zed Begins to lift his head
15 Ms. Jones Stay there, stay there, don’t get up
16 Jesse David, go to the sun!
17 David I tri-I am!
18 Ms. Jones So is there a flower-
19 Zed
20 Ms. Jones Nope, sit down. You’re dead, lay down.
21 Zed No, look it, no look it. I’m still alive!
22 Adam There is no other flower, Jesse!
23 Ms. Jones Why is there no other flower, Adam? So why, why isn’t there a flower up there?
24 Jesse There’s no information that, from the last time we did it, that there was a flower up there

Ms. Jones gives students the floor at the beginning as they offer suggestions and react to what is happening in the space. In turns 7 and 9 Ms. Jones uses a yes, and... move to agree with Zed’s declaration of death and urges him to “lay down” to show that he is a dead bee. Zed plays along with his teacher, yet when he realizes that the technology stopped tracking him (the Kinect cameras lost his image when he was on the floor), he stands up in turn 21 and declares that he is “still alive!” While at first Ms. Jones plays with Zed and tries to get him to lay back down, she realizes that he actually is being tracked once he stands up and according to the technology, that means he is still alive as a bee. She positions Zed’s return from the dead as no mistakes, only opportunities. Rather than erasing and replacing his idea, Ms. Jones allows Zed to resume his play from this new starting point. Instead of sticking to her plan, she laughs and turns her attention towards Adam and Jesse in lines 22-24 who are discussing ideas in line with the learning objectives, that the bees must translate the waggle dance from inside the hive to out in the field by starting their dances at the hive. As a result of this pivot, Ms. Jones skillfully puts the needs of the ensemble over the individual by getting back to the science content.

Figure 2. Map of how students initially translated their partner bees’ dances.

Ms. Jones’ turn to the ensemble is productive because it gets students talking together about how to translate the dance from the hive to the field. A few minutes into the discussion Ms. Jones realizes that in order to have a more focused conversation about the translation and continue to advance the narrative, she needs to direct students’ attention to where bees start from when foraging. She centers their conversation below (see Table 3).

Table 3: Episode 2, scene 2

<table>
<thead>
<tr>
<th>Turn</th>
<th>Speaker</th>
<th>Talk</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ms. Jones</td>
<td>Where do the bees always start, though, when we go out into the field?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Multiple</td>
<td>The hive!</td>
<td></td>
</tr>
</tbody>
</table>
Ms. Jones makes statements in turns 1 and 3 to help tell a cohesive story and focus students on curricular objectives. Her improv moves, along with the physical red arrow prop, help Zed connect how the dance starting point translates from the hive to the field. In the turns following, the students finally come to a collective conclusion that because the bees start at the hive, to find the flower you have to start doing the dance at the hive rather than an arbitrary starting point. Although our case is from a mixed reality environment, teachers in all sorts of situations experience this tension between structure and responsiveness in pursuit of telling a story, and thus instructional improvisation can be a way for all teachers to effectively respond to diverse student contributions during science inquiry.

**Discussion**

Instructional improvisation can help teachers think about balancing structures and responding to student needs. Furthermore, students can exercise agency and understand concepts even when the teacher acknowledges student contributions that fall outside of normative science. The rules—including agreement with contributions, yes, and- ing to elaborate on prior knowledge, making statements to advance learning, positioning mistakes as opportunities for learning, prioritizing the work of the ensemble over individuals, and telling a cohesive story about content—are flexible and allow for ambiguity. For example, although Ms. Jones takes up Zed’s untimely death, the ensemble discusses and constructs a cohesive story about lesson objectives that does not include tangential lesson digressions.

Although our case study is from a mixed reality environment, we think that instructional improvisation can apply to other contexts as a way of organizing learning and discourse in science more generally. Although improv does not describe all of the moves that make a great conversation or lesson, as a lens on interaction it can help us understand the value of certain key moves. The effects of instructional improvisation include students having agency and joint ownership over science knowledge—students came to conclusions on their own as Ms. Jones used improv to support their inquiry. Additionally, we believe our model can support productive interactions that complement current reform recommendations. For example, Michaels and O’Connor’s (2012) “talk moves” is a framework for productive science talk that details types of conversations students should be having with a focus on the academic purpose of the conversation and teacher moves that reinforce this purpose. Our model can further reinforce frames like “talk moves” by focusing the teacher on the nature of the joint inquiry and opening space for student agency in conversations. Instructional improvisation emphasizes the subjective experience of what the classroom conversation should feel like—a playful, spontaneous co-construction of science inquiry. Our model also complements ideas like Reiser’s (2013) of building coherent storylines in science. Yet while storylines emphasize how conversations ought to be structured in ways that build coherence for students, instructional improvisation focuses on the qualitative experience of the collective interaction of students and teachers working together to write the story through a collaborative, coordinated effort. Future work might include explicitly supporting instructional improvisation in classrooms to observe how teachers take up and implement their interpretations of the model in different contexts.

**References**


**Acknowledgements**

The authors would like to thank OpenPTrack, UCLA REMAP, Kelly Peters, and the STEP teams at University of California, Los Angeles and Indiana University. This work was funded by National Science Foundation grants IIS-1522945 and IIS-1323767.
Designing for Rightful Presence in 6th Grade Science: Community Ethnography as Pedagogy

Angela Calabrese Barton, Michigan State University, acb@msu.edu
Edna Tan, University of North Carolina at Greensboro, e_tan@uncg.edu

Abstract: To take up equity concerns seriously in teaching and learning STEM, the ways in which teaching and learning are historicized and relational activities need to be considered. The construct of rightful presence – and the co-constructed “making present” practices which give rise to moments of rightful presence – is one way to consider more systemically how to make sense of the historicized and relational nature of teaching and learning. Our study shows how integrating the equity-oriented design approach, community ethnography as pedagogy, as a part of teaching and learning STEM, supports the emergences of making present practices: modeling ethnographic data, re-performing injustices towards understanding and solidarity, and making change dynamic and visible in classroom work. These practices and their implications are described using ethnographic data from 6 urban middle school classrooms in two states during a STEM unit focused on engineering for sustainable communities.

Introduction

We built the Occupied because kids were getting walked in on in the bathroom. It’s terrible. It’s a big problem. . . Our project solves it because it shows everyone when the bathroom “is occupied” (uses his fingers to put the words in quotes). Now you can’t just walk in and pretend you didn't know. . . I think this is a good problem to solve. I think it will help our community. Especially boys. The girls usually have someone watch the door, but a lot of us don’t. Mateo, 6th grader

In this quote, Mateo, a 6th grader, describes an engineering design solution, “The Occupied,” he and two peers prototyped over the course of a six-week unit focused on engineering for sustainable communities. The Occupied is a lighting system that allows classroom members to know when the bathroom is occupied. In his school, each classroom has its own individual bathroom located in the corner of the room, but the bathroom does not lock. The Occupied has 3 gumdrop 10mm LED lights in parallel circuit affixed to the wall outside the bathroom door. It uses the bathroom light as a switch to activate the solar panel, which powers the LEDs, connected by 12 meters of copper tape. Getting to the point of a working prototype was not straightforward for the group. They needed to figure out how to make the design work from a science and engineering standpoint. They needed to garner support from their teacher, Mrs. J., who admitted to being unaware of the bathroom bullying problem, and was skeptical of the students’ abilities to succeed with a complex project.

How did the three students, their peers and teacher work towards new understandings and relationships which made the Occupied possible, and which shifted how the youth were positioned and valued within the classroom and school community? In this study, we build an argument for the importance of framing teaching and learning and its outcomes through a critical justice view of “rightful presence” – or legitimate membership in a classroom community because of who one is (not because of who one should be), where the practices of that community work towards and support re-structuring power dynamics towards more just ends through making both scales of injustice and social change visible.

Our two guiding questions are: 1) In what ways does community ethnography as pedagogy support new relationalities among students and teachers in the science classroom? 2) How do these relationalities challenge traditional knowledge and power structures in learning, working towards a rightful presence in science class for youth from non-dominant communities?

Theoretical framing: Critical justice and rightful presence

We ground our work in a dynamic view of learning, concerning ourselves with how people, ideas, tools, resources, bodies and relationships move as people engage in social practice towards new futures (Gutierrez, 2012). Focused on more than vertical movement (e.g., such as novice to expert), this dynamic view illuminates the hybridization and transformation of practice as people move horizontally, from place to place, widening what counts as expertise. This dynamic view, importantly, draws attention to equity-related concerns because it foregrounds the ways in which individual experiences of injustice intersect with systemic injustice through sanctioned power hierarchies and practices (Squire & Darling, 2013). Learning involves the process of re-authoring and re-mixing practices from a wide range of experiences, located in the home, community, and school among other places, towards re-orienting social relations and knowledge hierarchies.
The long-standing education debt (Ladson-Billings, 2006) has produced patterns of participation in science classrooms that are oppressive for youth from non-dominant communities (e.g., Who has power in classroom settings, how and why). Current reforms in science education are grounded in the belief that if all students have access to high quality expectations and instruction, then educational equity will be achieved. While supportive of good teaching, such reforms do not go far enough in addressing seemingly intractable equity challenges across the intersections of racial and class inequality as outlined in Ladson-Billings’ education debt. Until these concerns are addressed as a part of teaching, then equity may not be achieved.

New approaches to teaching and learning that are equitable and consequential are needed. We use both terms, equitable and consequential, because we seek to purposefully call attention to the importance of recognizing teaching and learning STEM as an historicized and relational practice (DiGiacomo & Gutiérrez, 2016). Expanding upon patterns of participation, and who and what is recognized within classrooms are necessary to disrupt participation boundaries and knowledge hierarchies (Gutiérrez, & Jurow, 2016). Questions around who is teaching or learning, what is taught and why, are located in sociohistorical context (Engeström & Sannino, 2010) and central to interrogating normative practices and power dynamics in classrooms.

To make sense of how to design towards equitably consequential teaching and learning, the construct of a “rightful presence” is helpful to consider as a part of teaching all students in science class to engage in STEM. “Rightful Presence” emerges from critical justice studies of guest-host relationships in sanctuary cities, troubling assumptions of what it means to be inclusive in-the-present, and not in some abstract future (Squire, 2009). We can think of a rightful presence in education as legitimate membership in a classroom community because of who one is (not because of who one should be), where the practices of that community work towards and support re-structuring power dynamics towards more just ends through making both scales of injustice and social change visible.

One way to think about the on-going enactment of rightful presence is through the practices of making present which support it. Making present practices are always a reflection of struggle and foreground relationality, linking places and time. The connections of places and time in practice is central to authoring a rightful presence because it helps to reconstruct place through the enactment of new practice, borne out of minor politics. Such belonging breaks down the binaries between outsider/insider and novice/expert, “not by pursuing inclusion into an already established order; rather, it seeks to assert a new measure of justice even if that means undoing the order we currently exist in and benefit from” (Vrasti & Dayal, 2016). A critical justice view of rightful presence is a powerful frame for understanding equity-oriented teaching efforts in school science. Power dynamics are always at play in classrooms, acknowledged or not. Students are guests in their classrooms, expected to follow routines with the threat of disciplinary sanctions for noncompliance. Youth historically marginalized in both science and schooling, even if positioned as a welcomed guest, are expected to reconfigure themselves towards the majority culture. Their participation is always marked as guest, subject to an uneven power dynamic. This is an equity concern for youth from non-dominant communities whose cultural knowledge and practice has historically been marginal to school and to STEM.

**Methods and context**

We have been working with teachers in schools that serve non-dominant communities on integrated science and engineering towards the goals of rightful presence. We sought to integrate community ethnography as pedagogy towards increasing the possibilities for enacting making present practices as a part of teaching and learning STEM. This work occurred during an integrated science and engineering unit on “how can I make my classroom more sustainable?” Students were engaged in learning about the disciplinary core ideas of energy transformations, sources and systems, while also exploring the practices of engineering design. As a culminating project, students were given the design challenge to use what they had been learning to develop a design project in support of classroom sustainability. The design challenge was bounded with the following criteria: Students had to innovate something in the classroom in a way that would address a classroom sustainability concern. They were required to use a renewable energy source, such as solar panels or handcrank generators, 10mm gumdrop LED lights, copper tape, and any materials readily available in their classroom.

Teachers integrated community ethnography as pedagogy in two ways. First, teachers incorporated community dialog throughout the design process, including observations, surveys and informal conversations. Second, teachers incorporated multiple feedback cycles with different community constituents, and coordinated these feedback sessions with different points in their design cycle.

We used critical ethnography with participatory research and design approaches (Cammarota, & Fine, 2010) towards collaboratively working towards new social futures (Gutiérrez & Jurow, 2016). We seek to understand how students engaged in “equitably consequential” learning in the context of middle school engineering and what tools and practices might support teachers in enacting equitably-consequential teaching.
Critical ethnography is rooted in the belief that exposing, critiquing, and transforming inequalities associated with social structures and labeling devices (i.e., gender, race, and class) are consequential and fundamental dimensions of research and analysis.

We co-designed our materials with youth and teachers using participatory methods, as an approach to “social change making” (Gutiérrez, 2012). Our research team developed the unit and then co-revised it with youth and teachers first in afterschool settings and then in a summer camp prior to classroom implementations.

Our data sources include: 1) video recordings of a subset of class sessions, 2) conversation groups that occurred at the end of the implementation phase, 3) field notes, 4) artifact interviews with subset of youth about their engineering design, 5) conversation groups with teachers before and after implementation, plus daily informal conversations, 6) the youths’ work, and 7) field notes. Data was analyzed in the grounded theory tradition, using a constant comparative approach (Strauss, & Corbin, 1998). The first phase of analysis involved open coding by thoroughly perusing all generated data to surface a) critical episodes of engagement in the engineering design work; b) the knowledge and practices that youth drew upon during critical episodes; and c) how they iteratively defined the problems they were seeking to solve. With the help of our theoretical framework (rightful presence), we then worked to make sense of teacher and student interactions especially around community ethnography and community knowledge; why youth took the actions that they did; and the meanings the artifacts youth produced had for them, individually and collectively. This axial phase of coding was used to uncover relationships and connections between the youths’ science and community knowledge and practice, and their efforts to solve problems with their knowledge/practice for themselves and their community.

**Teachers and schools**

Mrs. J has been teaching for 33 years, but only the last three at Wilkenson Road School. She cared for her students but viewed their struggles as something that worked against her academic and social goals. The students love her. As one student says, “she is an awesome teacher.” Wilkenson Road School is one of the most diverse schools in the city, with 32% white, 28% Latinx, 8% Asian, 22% Black, 9% two or more races, and 1% Native American. It was converted to a “STEM” school 4 years ago in an effort to stanch the flow of students from the district into the local charter system and other districts allowed by state policies. By District accountability policies, the school does not have a strong reputation for academic success, with only 11% annually meeting passing levels on state exams (compared with state average of 33%). However, the school community rallies around supporting and welcoming all of their students, such as hallway displays celebrating the diversity of its student body, strategic partnering of students, hosting culture nights, and teachers encouraging students to “help each other” often.

Ms. D has taught for 12 years at the same middle school in a midsized city in the South, Sage Middle School. She teaches 6th grade Science and Social Studies, and is beloved by her students. One student described Ms. D as, “da bomb dot com!” Sage Middle School serves a diverse population of students with 43% Black, 38% White, 11% Hispanic, 5% Biracial, 3% Asian, and less than 1% each Native American and Native Hawaiian. 58% of the students come from low-income families. The school also serves 21% of students with a range of disabilities. Led by a dynamic school Principal in his second year, Sage won the “Most Improved Middle School” award in the district for most improved test scores. While the school displays overt signs of friendship for all, incidences of bullying regularly occur. The school also reported a disproportionate data of disciplining African American boys over all students.

**Findings and discussion**

**Community ethnography supporting Making Present practices**

We draw from two case studies to present our claims: The Occupied (described in the introduction) and the Happy Box. The Happy Box, designed by three girls at Sage MS during the course of their integrated science and engineering unit, is a light up messaging system that reminded all students that they were welcomed and valued in their classroom, and that social issues that concern students (beyond test scores and traditionally-valued achievements) are relevant. It is powered by two solar panels that lights up a design on the cover of the box with 5 LED lights. The design is of children around a globe, including a transgender child. The Happy Box functions as a classroom mailbox with individualized envelopes for each of the 28 students in the classroom. Teachers and students can use the Happy Box to write encouraging notes. The Happy Box is checked right after morning announcements by the Principal, which typically follows the script detailing upcoming exams, test scores, and school achievements.

We selected these two cases because they reflect different ways of thinking about and responding to classroom sustainability, and come from two different classrooms in two different states. We describe three forms of making present practices that emerged as community ethnography as pedagogy was interwoven into the
teaching of integrated science and engineering: 1) Modeling ethnographic data, 2) Re-forming injustices towards understanding and solidarity, and 3) Making social change dynamic and visible. We describe how these practices emerged, and show how they worked towards moments that allowed for new relationalities to emerge – restructuring positional and knowledge hierarchies among community, students, teacher and science – conditions, we later argue, necessary for moments of rightful presence to emerge.

Modeling ethnographic data
Community ethnography was used throughout the design challenge in numerous ways. As youth sought to determine what problems to solve, they conducted surveys and interviews with school peers and staff and community members focused on the sustainability concerns community members faced, and what those concerns meant to them and why. These data became important resources the youth drew upon to organize their experiences in their classrooms and community. They graphed survey results and conducted content analysis of comments made by the people they surveyed and interviewed. Modeling their ethnographic data in these ways served as a powerful making present practice for these representations made visible both the classroom or school norms and routines which the students found oppressive and they helped to illustrate scales at which they occurred. These representations also positioned the students’ experiences as valuable scientific data.

With the Happy Box, the youth used graphical analysis of ethnographic data to make present concerns about bullying and lack of fun as current and salient issues to be tackled in science class projects. As Kristen explained: “From the survey we found that 55% of the community members felt that school was too stressful and not fun. 45% also said that we do not celebrate and encourage the community enough, we only celebrate grades. We decided to make a Happy Box where encouraging messages and cards can be left for classmates . . . and know that people care.” The power of the numbers, and their visibility in graphed data, legitimized their experiences feeling marginalized by this testing-oriented culture. The girls also surveyed schoolmates whom they did not typically come into contact with. Learning that they shared stark similarities with their own concerns about their school further affirmed their wish to address this concern.

Kristen’s transformation as a result of the community survey was significant. Ms. D was “amazed” at how Kristen became “totally into the project after the survey.” She said: “She made me sit with her during lunch and showed me what she and her group are going to design. The Happy Box. She ran it. She was the leader.” Ms. D agreed with Kristen’s group that “it is really hard for kids to transition from whatever turmoil they just endured in their home lives or out of school…and you wouldn’t believe what some of these kids have to face each day…and to come to school and be expected to immediately switch-on and be focused.

With the Occupied, the youth conducted content analysis of open-ended survey items and interviews with community members. These detailed comments offered youth the language to describe their experiences in ways that illustrated how the problems the community faced were much bigger than themselves. When Tryne reported that one of their survey participants indicated that “people were worried about people just barging in our classroom” the group had a way to frame their own experiences with the bathroom.

In both examples, the data provided the language, the visual representation, and the evidence to allow the students to name and push back against the dominant narrative that there were no real injustices that were solvable in the classroom. They also provided a way for the youth to connect problems within classrooms to broader problems within the school and community. Ms. D and Mrs. J were surprised at the social issues revealed by these surveys that they were not aware of and supporting students in survey content analysis helped give the students the words to name those problems to solve in science class.

Re-performing injustices towards understanding and solidarity
As students imagined projects that might solve some of the problems they identified, they used their ethnographic data to re-perform the injustice documented. We see this idea of re-performing injustices towards understanding and solidarity as a second making present practice. This practice was particularly powerful in how it involved both teachers and students using hybrid discourses to justify engineering design considerations, such oral testimonies and narratives of personal experience.

As the Occupied group sought to explain how bathroom bullying worked so that they could attend to technical design features, they re-enacted narratives of their experiences of being barged-in upon. In one re-enactment, Sophia walked into the bathroom while Mateo knocked and put his ear to the door. As he did this he narrated: “See, the rule is that we have to knock on the door, and put our ear to the door to hear if anyone is in there. If we don’t hear anything its clear to go in. But, some kids pretend they don’t hear nothing, and they walk in.” He then walked in on Sophia, who re-enacted the humiliation of being walked in upon.

These re-enactments served to connect their projects to the broader group of students in the class as they garnered attention with their dramatics. Such re-performances appeared to crystalize moments of refusal to be
The Occupied had produced a sketch-up of their design, which described the main technical specifications: a 3v solar panel that connected to a single white LED light, connected by 12m of copper tape with a layer of protective electrical tape to guard against wear and tear. The solar panel would be activated by turning on the light in the bathroom. The sketch-up also included one main social specification: To make the light so that people would see that the bathroom was occupied. When sharing their sketch-up with community members to solicit feedback, each youth, offered corroborating testimony about how the bigger (and hidden) problem of bathroom bullying were the rumors that bathroom bullies start in school hallways about what they purported to witness. These rumors, Mateo reported, were what “really hurt” because they were not true. While the Occupied in some ways is humorous, it is also serious. For Mateo, being "walked in on" while using the bathroom represented a safety issue that he and his friends had experienced. This experience had spread out with students’ rumors outside the classroom making students to "never [go] to the bathroom during the day anymore."

As the students were encouraged to solicit community ideas as homework, Mateo started bringing in electrician supplies from home, further building on the connections to cultural knowledge and practice that the ethnographic activities opened up. He started role-playing the master electrician, wearing his uncle’s electrician shirt, he brought in electrical tape, and began to tell stories of learning to build circuits from the age of 3, with gummy worms as he went on the job with his uncle. His peers enjoyed his stories and they commented on how cool his shirt was. These re-performances, which took the form of dramatic re-enactments, role playing, and testimony, legitimized students’ experiences of systemic oppressions as a part of engineering design, connecting the power of their experience with what it meant to engage in STEM.

With the Happy Box, such re-performances took shape through the illustrations the girls included in their Box. In their initial sketch-up design, the girls intended to construct two simple circuits, each lighting one LED light, to decorate the front side of the box. They thought that coin-cell batteries might power one LED light for a fairly long period of time before they would have to replace the batteries. However, when choosing a box to work from out of all the recycled boxes available, the girls chose a children’s shoe box which had a whimsical design on the box cover. Kristen and Elsa really liked the design and proclaimed it “Happy!” They were particularly drawn to the globe image with children surrounding it in a circle.

Here, the girls sought to re-enact their desire to be happy, and to help others to be happy, by situating it within a playful design. As the design included children around the globe, they sought to further connect their desires to be happy with children globally, signal a concern much bigger than their classroom. They liked that the “children around the earth means everyone is important.” Later, the girls spent time decorating the front of the Happy Box by choosing which of the patterns to color in and highlight, to draw attention to specific features. Elsa pointed out: “See these children around the globe? There are 11 of them. We colored 5 pink for the girls, and 5 blue for the boys, and the one right on top is half pink and half blue, for transgender kids, because you gotta include everybody. Everyone is important. We want to light up the kids, maybe a few of the boys and girls but definitely the transgender kid.”

As this quote indicates, re-performances made visible the struggle of young people to belong in their classroom schools and community. They not only linked places, such as classrooms, with local and state politics, they put in productive tension the oppressions the students sought to flee and their refusal to be victimized. With the Happy Box, the girls specifically gave witness to emotional needs of transgender children. Given that their home state of NC had been embroiled in a bathroom controversy regarding transgender children and bathroom use, this is not a trivial concern. With the Occupied, re-enactments and testimonials allowed the silent bullying to be made problematic – to be laughed at as it lost its power over the youth.

**Making social change dynamic and visible through engineering**

On-going dialogue among students and community members supported critical considerations on whether and how emerging designs addressed the concerns that students cared about. These on-going dialogues gave rise to a third making present practice: Making social change dynamic and visible through engineering/making. This was an on-going and iterative practice that not only opened opportunities for the youths’ experiences of historicized injustice to be challenged in the classroom, but it also worked to position the youth as highly capable and creative with their STEM/Making knowledge.

When community members visited to provide feedback to the construction of designs, the Occupied learned that some were concerned that the one LED light would not be visible enough to alert classroom members that the bathroom was occupied. The group decided to add 2 additional lights (forcing a move from a simple to a parallel circuit) and black construction paper as a background. This required the group to test multiple solar panel styles to find one that generated enough voltage for 3 LEDs, and that would not melt when placed so close to the incandescent lightbulb in the bathroom.
Mrs. J fretted that the Occupied would not get their project to work given the new layers of complexity. She admitted to not quite knowing herself, from examining their technical sketch-up and design, whether it could work or how to help them fix it. But Mateo overheard and challenged her, as Mrs. J describes: “I truly did not think they could pull that off. Never saw them frustrated. They just went right back to work. In fact [Mateo] after he heard me say that said to me, Be honest, did you really think we could not do this? I said in all honesty…I didn’t think you could pull it off. I am impressed. He just smiled. He can be difficult but he handled this with a smile almost the whole time. And confidence.” This appeared to be an important moment when Mrs. J eventually recognized what the group could accomplish and what they cared about. She told them that the project would “help the class so much.”

The youth explored and experimented with different technical options so as to not dilute their social message: Three brightly lit LED bulbs would literally send a bright message about the fact that the bathroom is in use, and no one should barge in. Leveraging on complex engineering practices to send a strong social message was also evident in the Happy Box case.

The Happy Box group wanted to send a message that all students matter so they made sure that girls, boys and transgender students on their Box were lit up. This required them to experiment with different circuit types so that they could light five LEDs with one hand-crank generator. However, through community dialogue, the girls decided the lights needed to stay lit when messages were present to remind people “everyone matters in our class all the time” and to not disrupt classroom routines, when a student needed to stand up to crank the generator to power the LED lights. The girls had to change their power source to solar because a handcrank generator, without a capacitor, would not keep lights lit. This presented the further problem of needing to split the lights into two parallel circuits due to the power demands of 5 lights and smaller solar panels.

They had to rethink their circuitry and energy source, because they “are not going to use five batteries and keep replacing them, it’s expensive.” Julia tried out three different versions of parallel circuits in a circular form. After building three differently shaped parallel circuits because of the positioning of the lights (which had to light up the ‘heads’ of the figures), the group managed to get 3 of the five lights to light up with a hand-crank. Julia was determined to get all 5 lights to light up and continued to mull over the circuitry with Edna’s suggestions. 5 LED lights, with the transgender child’s LED in the center, was critical to the girls. The group decided that the LED lights had to be lit continually to show that “everyone matters in our class all the time.” The class had access to small solar panels, and one would not be enough to power 5 LED lights. After more collaborative work, they settled on the two parallel circuits, one powering 3 lights, the other with the remaining two, and each connected to a flexible solar panel that is taped on the outside of the box.

This practice directly linked engaging in dialogue with community members with opportunities to deepen STEM expertise and to be recognized for it. The process of constructing and reconstructing designs as a result of new insights from community dialogue became physical manifestations of historicized injustice and a refusal of victimization, while also reflecting strong connections between classroom, school, and community. These are not minor changes for sixth grade students, but the girls worked through the complexities so that their design attended strongly to the problems they identified.

How did these practices support a rightful presence in science class?

Making present practices challenged the normative culture of learning and participation in their science classrooms, creating conditions that made moments of rightful presence possible. They also helped to make visible the ways in which intersecting scales of injustice play out in science classrooms, denying students a rightful present there. We believe that as making present practices were continually re-enacted and re-created through a range of localized activities and informal encounters, rightful presence emerged, at least in moments. The data suggest that are two intersecting ways in which to think about rightful presence.

First, making present practices restructured positional and knowledge hierarchies allowing moments of rightful presence for individual students, previously marginalized, to emerge. We are concerned with how, and by whom, youth were recognized and valued for their expertise, reifying rightful presence. For example, the making present practices enacted as part of constructing the Occupied provided opportunities for students like Mateo to build upon and share the knowledge and practices he brought from home. But, Mateo, like many of his peers of color, has been positioned as an outsider to school science – a young boy whose expertise from home and success in his project surprised his teacher. And yet, Mrs. J. cared about Mateo as she cared for many of her students. She worried about his home situation, where he was being raised primarily by his uncle while his father was incarcerated. She often tried to get him to join group meetings and help them with the projects, but he would not participate. She often called him out to stay in his seat, when he roamed around the classroom during instructional time.

During the construction process and as the groups sought to iteratively refine their projects in response to community feedback, Mateo and his groupmates had to work repeatedly from their desks on the far side of the
room to the bathroom to work on their project. As Mateo moved about, Mrs. J. noticed that some of the students began to ask Mateo for help when their circuits were not working. She began to position him as an expert, especially when she got stuck, “It’s wonderful for kids too. I can say, ‘I don’t get this. Mateo could you look at this.’” Sometimes you can find student experts.” Mateo was recognized for his expertise, and his freedom of movement opened up in the classroom. This is not minor. He is a student whom Mrs. J initially described as a troublemaker with a “sad” background. Mateo’s emerging identity as an expert in circuits gained a rightful presence in the classroom, overshadowing his “sad student” identity that Mrs. J has thus far held as Mateo’s sole identifier. Moving as an engineer across the classroom to work on his “site” (bathroom) further solidified Mateo’s rightful presence as a classroom electrical engineer.

Likewise, engaging in the community ethnography provided Kristen, Elsa and Julia to establish a rightful presence in middle school engineering, where their concerns about the school’s punitive culture that students suffer from became the impetus to engineer. We see how the girls further establish their rightful presence with their Happy Box innovation. Thus, making present practices help students bring to bear issues salient to both their everyday lives at school and their science learning, issues that may not be immediately deemed relevant, or even known to exist, by their science teachers.

Second, making present practices made legitimate in classroom culture particular relationalities previously unsanctioned, while creating tangible symbols of rightful presence. Both projects became a part of classroom practice. Each time the bathroom was used in Mrs. J’s classroom, the Occupied lit up, a visible reminder of Mateo and his groups’ innovation and expertise, and the class had the opportunity to collectively monitor the bullying situation. About a month after the unit was completed, one of us stopped in the classroom to talk with the teacher and stayed to help some students with their work. A student had gone to use the bathroom. Another student tried to barge in on her, at which point, several students hollered in unison “the lights are on!” at the barging student. When 5th grade graders visited Mrs. J’s classroom, both visiting teachers and students pointed to the design as important and something they needed for their own bathroom problems. Soon, many teachers were requesting the same system be installed in their classroom. Here we see how these making present practices legitimizing “change” as an important outcome of STEM learning.

Discussion
We have argued that to seriously take up equity concerns in teaching and learning STEM and STEM-rich making, we need to consider the ways in which teaching is both an historicized and relational activity. The construct of rightful presence is one way to consider more systemically how to make sense of the historicized and relational nature of teaching. Whether and how students are recognized and valued for what they bring to learning as well as how they are supported in more expansive outcomes of learning all are shaped by – and shape – the extent to which one has a rightful presence in science classrooms.

Our study shows how designing for community ethnography as a part of engaging in STEM-rich making in the classroom support the emergences of three making present practices. These practices opened up new modes of previously unsanctioned relationality among students, teachers, community and disciplinary knowledge and practice. Students had new opportunities and structures for being recognized for their experience and legitimized spaces for doing so. Students were not welcomed as fuller members of the science classroom simply because their science/engineering expertise grew. Other forms of expertise also became important levers in locally important ways. Having multiple forms of expertise and ways to enact it towards solving injustices were the process and product of science learning, a vastly different scenario than in most science classrooms.

Using a rightful presence perspective to lens our work in the classroom settings also revealed distinct differences from its previous conceptualization. First, the making practices identified in the borderland and refugee/immigrant literature are public and shared by those willing to take them up. What we see as powerful in our study is that the making present practices enacted by youth were done so directionally; that is, in part, in solidarity with their teacher, a specific point of power in their classroom, and a potentially important ally who students had. The youth explicitly called attention to intersecting scales of injustices in their work, in terms of school, class and science, at both the local and sociohistorical level (Cole, Kaptelinin, Nardi, & Vadeboncoeur, 2016). The Happy Box had rippling scales of impact, from the immediate (individual students who writes and receives notes) to Ms. D’s classroom culture, to the school culture, similarly with Occupied. Further, projects like the Occupied made visible the immediate problem of bullying in the classroom, and attention to its sociohistorical location as it primarily affected and further marginalized boys of color. One aspect is that the relationality sought for by students reflect
the institution of public schooling with oppressive structures, manifested in practices that value and normalize ideologies of test scores, strict discipline, racialized and gendered experiences.

Third, making present practices expanded social networks, increasing more opportunities to broker for rightful presence. Youth and teachers strategically brought new and diverse people into the design conversation, such as their friends, parents/grandparents, teachers, engineers, and also little kids, incorporating the technical and social concerns discussed into their designs. This allowed the youth to advance the technical quality of their innovations while deeply enouncing themselves as an integral part of their design. While refugees physically flee places of oppression in hope of safety and acceptance in foreign sanctuary cites, the students here are bound, literally and figuratively, to reinhabit their physical school spaces through struggling for rightful presence, through moment-to-moment efforts (Gutierrez, 2016).

Making present practices, while potentially empowering, however, are fraught with risk. For many students, especially students of color and students in poverty, a rightful presence in the science classroom is not guaranteed. The current structures of schooling and science actively work against it. The design tools of community ethnography supported the production of making present practices in ways that open up moments of rightful presence. These moments are often transient. Perhaps, optimistically, we can argue that such visibility opens up moments of rightful presence, which can and must be built upon. The emergence of making present practices support the youth up-ending standard expectations of what it meant to make, to produce an engineering design and/or a school artifact. As the histories and geographies of the youth shaped the ways in which they defined the problems and the solutions they developed, the youth disrupt the historically established notion of what counts as STEM in school settings and whose knowledge or practices matter in STEM.

Conclusion
The dominant equity narrative in STEM education is problematic because it does not align with the goals of justice. It positions youth from non-dominant communities as inferior and in need of remediation if or when they “lack” how the field has framed what counts as STEM. The youth in these classrooms, through enacting making present practices, pushed back against normative structures in the science classroom. They engaged in design work that leveraged what they learned in their class, bent towards justice but also unpredictable ends, but ends that also opened dialogue around the problems they faced and their capabilities in responding to them.

References

Acknowledgements
This work was supported by the National Science Foundation DRL #1502755.
Designing Technology as a Cultural Broker for Young Children: Challenges and Opportunities

Yanghee Kim, Northern Illinois University, ykim9@niu.edu
Sherry Marx, Utah State University, sherry.marx@usu.edu
Hung Pham, The University of Waterloo, hv.pham.2704@gmail.com
Tung T. Nguyen, Auburn University, tungisu@gmail.com

Abstract: This study developed a socio-technical learning community of a humanoid robot, one child from a native-English-speaking background, and one child from a Spanish-speaking background, both living in the U.S. Grounded in pedagogical and communication theories, the bilingual robot mediated the two children’s interaction activities to invite both children to participate equitably. Core research questions included i) What are the challenges in designing a robot to mediate equitable, collaborative interactions among young children? and ii) What themes arise in children’s interactions with the robot and each other? We adopted a design research approach to developing interaction episodes and the robot’s mediating utterances, where our designs were continuously revised as we observed triadic interactions in a natural kindergarten setting. This paper discusses our design experiences, as well as themes that emerged from ethnographic observations over a four-month period.

The major issue addressed
Recent National Assessment of Educational Progress test results in the U.S. indicate an achievement gap and a very flat trajectory for lower-performing students, especially language and cultural minority students (McFarland, et al., 2017). More problematically, deficit thinking and marginalization prevalent in the classroom have taken a toll on both the learning and the identity of these students. Minority children are often viewed by educators and classmates as having deficits in language and culture that prevent them from successfully contributing to the classroom community (Valencia, 2010). Over time, children can adopt these messages and learn to identify themselves with marginalized communities. This identification with marginalization can start as early as preschool and become more entrenched as children grow older (Van Ausdale & Feagin, 2001). Once they develop a negative learner identity, children are less likely to recognize mainstream paths to success through schooling (Nasir, 2002). The high dropout rates for minority students in U.S. schools are tied to this marginalization from the mainstream learning communities of schools and classrooms (Gándara, 2010). Even though many students reach proficiency levels in English language as they move through their school years, their academic achievement often times does not improve. Rather, the dropout rates of language minority students increase as they age (Boone, 2013).

In reality, children coming from diverse backgrounds can enrich the mainstream school culture with their unique cultural and linguistic assets if they are only given the chance to do so (Vasquez et al., 2011). It is very warranted to provide an inclusive learning community, where every student is valued and welcome. In such a community, all children are encouraged to build on their prior experiences (Donovan & Bransford, 2005) to participate actively and make progress toward academic success. While coordinating educator training and efforts to overcome unconstructive deficit thinking is imperative, such processes can be quite time consuming and difficult to achieve as educator beliefs about children are often subconscious and difficult to change (Borg, 2009). As an alternative, the authors explore ways to provide an inclusive and equitable learning community for diverse children quickly, with the help of an unbiased, embodied technology, in this case a humanoid robot.

In our research introduced in this paper, we sought to develop a socio-technical, inclusive learning community of a robot and children, where the robot might facilitate equitable collaborative interactions among children coming from different backgrounds. We first developed the theoretical model for robot-mediated collaborative interactions, grounded in theories of child development, multicultural education, and intercultural communication. This model was implemented in human-mediated and robot-mediated interaction activities sequentially with kindergarten children. The robot’s mediating utterances and the interaction activities were refined in an iterative cycle as we ethnographically observed the interactive sessions in a classroom setting.

Potential significance of the work
As the student population becomes increasingly diverse worldwide, designing inclusive school learning environments that embrace linguistic and cultural diversity has been a constant challenge. Also, being able to collaborate and appreciate differences are essential skills children need to master as they grow inside and outside
schools. However, these skills are not always addressed in regular school curricula. This study explored using a humanoid robot to help close this gap, creating an inclusive socio-technical community where children learn to work together equitably regardless of their backgrounds. As we observed children’s interactions with each other and with the robot in a natural setting at their school, we were better able to identify children’s needs and learn what worked well and what was lacking in our designs. Here, we discuss our design challenges and lessons learned, which could be useful to other researchers designing for young children.

Research in learning sciences has traditionally treated cognition and affect as distinct constructs and explained the processes and roles of cognition and emotions for learning separately. However, recent neuroscientific research informs us that our thinking, feeling, and context are by nature invariably intertwined (Immordino-Yang, 2016). That is, children’s emotions and cultural contexts serve as an inseparable rudder to steer their learning and intellectual development. In designing programs, therefore, learners’ holistic experiences as intellectual, social, and cultural beings should be taken into account in order to bring successful learning in the long term. Likewise, the development of positive learner identities through positive learning experiences is equally as important as academic skill development since positive identity is foundational for persistence in learning difficult topics, resilience to failure, and academic success. Our design approach to a supportive socio-technical learning community aims to reinforce positive social and emotional experiences as children develop academically.

Technologically, we have a long way to go to be able to implement natural dialogue between a robot and children. In general, research on designing for young children has not been as popular as designing for upper age groups. Speech and voice technology, particularly, is quite limited in recognizing young children’s speech. There is a great need for children’s interaction data to help build analytic models to advance this area of research. A corpus of speech data sets generated by this study can be a resource for researchers in natural language processing, who are interested in designing tools for young children.

Theoretical Background

Developmentally appropriate, multicultural pedagogy

According to child development research, kindergarten-aged children improve in fine and gross motor skills (Radesky, Schumacher, & Zuckerman, 2015) and like to engage in fantasy play (Giménez-Dasi, Pons, & Bender, 2014). They are rarely able to sit quietly for long periods and like to spend much of their time with peers. They become aware of themselves in relation to peers and begin comparing their performance to that of their peers, recognizing that the needs of others are often different from their own. Their family and cultural backgrounds have a great influence on their developmental characteristics. Not surprisingly, large individual differences are observed in motor agility, temperament, sociability, and academic performance among kindergarteners. For this age group, therefore, developmentally appropriate pedagogy may involve i) accommodating diversity in interests, background knowledge, and talents; ii) allowing children’s play and autonomy; iii) encouraging children to explore fantasy worlds; iv) providing the opportunity to practice new skills; and v) guiding children in ways to successfully interact with peers (e.g., in resolving conflict and playing in collaboration).

Multicultural pedagogy acknowledges children as cultural beings and fosters an egalitarian view of diverse languages and cultures (Paris, 2012; Ladson-Billings, 2009). This pedagogy seeks to make use of children’s prior linguistic and cultural heritage in the design of curricular materials and learning activities. In a culturally-sustaining learning community, a child’s home language and culture are respected as assets rather than deficits. These assets, or funds of knowledge (González, Moll, & Amanti, 2009), help children maintain a positive identity and transfer knowledge and skills from home to school (Moll, Amanti, Neff, & Gonzalez, 1992). Children are invited to share their cultural experiences and have an opportunity to become fully-engaged participants in the design of learning activities. In such a supportive community, all children may develop intellectually, socially, and culturally in an equitable way.

Intercultural communication

Communication is a process through which individuals or groups share information to develop understanding of each other and the world in which they live. Effective communication involves not only explicit verbal and nonverbal exchanges but also the interlocutors’ social, affective, and cultural characteristics, largely influenced by the context where the communication takes place (Carter & Fuller, 2016).

Identity is not only personal but also social and cultural since how we view ourselves is moulded through interaction with others (Harré & Moghaddam, 2003). When involved in a dialogue, we maintain our own unique sense of individual identity and build a common base of understanding (Bakhtin, 1987). By telling our personal
stories, we get closer, bond, and disclose things about ourselves. Empathy and listening with unconditional positive regard for one another are key to meaningful communication since these actions create a supportive psychological climate where the interlocutors will be willing to tell their stories (Littlejohn & Foss, 2011).

An opportunity to participate repeatedly in communicative contexts with empathy and positive regard are especially important for interlocutors coming from different cultures. Newcomers to a community learn the meanings shared in the community and participate in communal conversation, through which they negotiate between individual self-concept and community membership. Through prolonged exposure to the new culture, newcomers come to transcend their original culture and gradually build up new cultural schemas. Cultural schemas are sets of knowledge about appropriate behaviors and roles in specific situations in a particular culture. They are created from repeated participation in interactions with people who share common cultures in the same situation (Nishida, 2005).

A model of cultural mediation
From the review of educational and communication literature, we have derived three core approaches to the robot's cultural brokering actions: invitation, opportunity, and empathy. Invitation is necessary to welcome children into a learning community where they will be positioned as contributing, integral members. Opportunity is a set of circumstances that is frequently under-supplied in many formal education settings. With the robot, children will be given ample opportunity to initiate their interactions, practice conversation topics repeatedly, and participate in creative, challenging activities. Empathy requires that children be treated with respect and understanding; it is closely linked to relationship building that supports social and intellectual growth (Gudykunst, 2005; Littlejohn & Foss, 2011).

In addition, robot mediation aims to help children achieve three communicative goals (building common ground, building an equitable partnership, and building a co-cultural schema), which offer optimal conditions for equitable, inter-cultural communication. The first step for enabling children to work together is building common ground. Children need to feel comfortable with each other and share their personal stories in order to establish a minimum level of common experience and trust. Equitable partnerships emphasize that another’s autonomy and identity are as important as one’s own. This respect is developed through careful listening, openness to new experience, and collaborative interactions. Cultural schemas are built up through repetitive experiences in cultural situations. While they engage in interactive, imaginative activities in the robot-mediated learning community, children co-construct meaning, understanding, and identity in the unique activities they share.

Guided by this mediation model, we have instantiated a socio-technical interactive triad of two culturally and linguistically diverse children and a bilingual robot as an interaction mediator. Our design research took a grounded-theory approach and started with two foundational questions: i) What are the challenges in designing a robot to mediate equitable, collaborative interactions among young children? ii) What themes arise in children's interactions with the robot and each other?

Methodological approaches pursued
Over a span of one semester, we conducted design research, where we crafted initial designs for triadic interactions and refined them as we reflected upon our own designs and ethnographically observed children’s reactions to the robot and their interactions with each other in a natural kindergarten setting. For given the technological limitations in implementing natural interactions between the robot and children, we employed a Wizard of Oz method (Rick, 2012), where a researcher, hidden behind the scene, controlled the robot’s utterances responsive to children’s talks.

The robot system
The robot Skusie, robot controller, main controller, and server. Skusie is combined with a mobile phone and controlled by Android apps via Bluetooth technology. The phone is cradled on the robot’s head, acting as the robot’s visible brain. The body is equipped with sensors and mobility, accompanied by a wand with an embedded optical sensor and microphone. We employed a voice synthesizer that allows Skusie to speak in both English and Spanish. In the interaction sessions with children, the researcher (acting as a wizard and controlling the Main Controller) can manually provide speech utterances for Skusie, or select them from canned utterances in the interaction scenarios. The researcher can also control the Skusie’s motions.

Participants and context
Participants were twenty-four kindergarten children in a public elementary school in a mountain-west state of the United States. The school has a high rate of families living near or below the poverty line. School children were predominantly white English-speaking and Latino Spanish- and English-speaking. All participants were identified
as low performing by the school and attended a supplemental class that provided additional practice with language and academic skills for an hour around lunchtime. For the study, children were divided into twelve pairs, with an intent to form cross-cultural, cross-linguistic (English and Spanish) partnerships. The number of two language groups were not balanced. While all children participated in the interactive activities, the research team studied nine culturally diverse pairs.

**Design of interaction episodes**

In the design of interaction episodes, we applied the aforementioned developmentally appropriate and multicultural communicative approaches. We personified the robot, Skusie, as a new friend who just arrived from another planet (an imaginary world) and did not know much about life on earth. In this learning community, Skusie needed children's help in order to learn about everything including human language and culture, through which we sought to stimulate children’s sense of agency and autonomy as they helped the robot. Skusie spoke both Spanish and English but its speech in either language was not always perfect, just like the children. So everyone here was not judging but understanding. Children were asked to work together to teach Skusie, and they made use of their own knowledge and interest. Through this context, children were invited to help, given an opportunity to participate using what they already knew, and not judged but appreciated by their contribution. For the topics of interaction, the research team reviewed children's books in schools and libraries and chose four very popular topics for conversation: animals, birthdays, school, and family.

Referring to our mediation model, the design team drafted mediating utterances, which were used by a bilingual research assistant who acted as the mediator for the first six weeks. Based on these human-mediated sessions, we crafted the robot’s utterances and flows (called scenarios) for each episode, which were deployed later in robot/children triads. Simultaneously, the developers worked on software and hardware systems for robot-mediated sessions. During robot-children interaction sessions, another bilingual research assistant sat behind Skusie as a moderator to clarify instructions or intervene for smooth flow when necessary. When her intervention was needed, we recorded her utterances and added to the robot's utterances in the following sessions. By doing so, the moderator's intervention became minimal and later not necessary at all as robot interactions improved. Weekly, the research team met in full to review digital recordings and assess the strengths and weaknesses of the tested episodes and robot functioning. Improvements were suggested, honed, and then deployed the following interaction sessions. Overall, the robots' utterances and the flows for each of the four topics were drafted, tested, and refined over the entire four-month period.

**Data collection and analysis**

Researchers and assistants visited the same, supplemental kindergarten class two days a week from mid-February through mid-May, 2017. For the first six weeks, a bilingual research assistant acted as the cultural broker after she was educated about the robot mediation approaches. Adopting the approaches, the research assistant led a 15-minute activity with each pair of children. Children were presented a conversation topic and encouraged to interact with one another to engage with the topic through a loosely structured flow. Being bilingual, the research assistant was able to adjust the activity as necessary. During the second six weeks, we deployed the robot to interact with the pairs of children, using a Wizard of Oz technique. In this method, a researcher acting as a wizard controlled the robot while sitting unobtrusively in the corner of the room. The wizard controlled the timing and content of the robot's utterances. The robot met with pairs of children on the floor of a media center in the school. All interaction sessions were video-recorded and later typed into English- and Spanish-language transcriptions. A researcher also took ethnographic field notes of each activity, recording them in a researcher journal.

Forty-three 15-minute sessions with a research assistant or a robot interacting with pairs of children were digitally recorded and then transformed into detailed typed transcriptions. These transcriptions were analyzed in concert with the researcher’s journal, which included field notes from all classroom interactions and weekly research team meetings. Given the iterative nature of the design processes and the constant improvement of the scenarios, researchers looked for evidence of improvement in the four tested scenarios, using the framework of building common ground, building equitable partnerships, and building a co-cultural schema as markers of high quality interactions. Additionally, researchers used traditional ethnographic methods to constantly compare phenomena that occurred across scenarios and children to ascertain additional findings.

**Findings and implications**

**Design challenges and our solutions**

Our first question asked, *What are the challenges in designing a robot to mediate young children's equitable, collaborative interactions?* As we reviewed the market and the literature on developing advanced technologies
for young children, we identified four issues that might challenge the implementation of robot meditation. First, compared to adults, kindergarteners are still developing their language skills so often use words that approximate the meaning they intend. Their syntax is still developing; their pronunciation is often not clear. Speech-recognition software that can reliably understand kindergarten language does not yet exist. To resolve this, we used a human controller who acted as a wizard behind the scene to talk to the children through the robot. The strength of this arrangement was that the controller could hear what children said and input an appropriate reply. Limitations included an occasional delay between controller’s input and the robot’s utterances. This often resulted in Skusie not responding for several seconds, and then responding with too many utterances at once, interrupting the children’s interactions. While some children simply laughed at Skusie’s “hiccups”, other, shier children often became quiet as the following example of two girls, one Latina and one white, illustrates.

ROBOT: Tell me more about animals. What do you do with animals?
GWAV: [Starts to say something.]
ROBOT: Explicame mas sobre los animales. (Tell me more about animals.)
GWAV: [Starts to say something again.]
ROBOT: Que haces con los animales? (What do you do with animals?)

Our second focus was to design a robot that could speak in a kind, casual, yet direct manner, repeatedly inviting them to talk. To do this, we first had Skusie greet children with their names. Hearing their own names from Skusie was disarming and also engaging for children. At first, they could not believe the robot was talking directly to them. After repeated interactions, many children responded to Skusie naturally.

ROBOT: Hello BLED and BWLA. Good to see you again.
BLED and BWLA: [Laugh and sit down.]
[robot moves closer to them.]
BLED: Uh oh.
ROBOT: Hello BLED and BWLA.
BLED and BWLA: [Laugh]

Designing a conversation flow that was appropriate for children took some experimentation. Designers added questions to Skusie’s utterances so it could give children an opportunity to speak more: questions included, “Why?” “Why not?” and “Tell me more.” We also added the statement, “I’m confused,” which successfully got children back on topic if they digressed or spoke in a manner the wizard could not understand. Tricky was some exclamation sounds, such as “aww” or “ah,” which were meant to convey understanding with emotions, fell flat when the text-to-speech engine in the robot pronounced them phonetically. As this happened, we removed such sounds from the robot’s utterances.

Our third design focus was triggered by the fact that children were divergent thinkers and actors, and their responses were very often unpredictable. First of all, we did not want to limit their thinking, rather we imagined what children would do or say in a particular event, but our imaginations were limited by our own adult experiences. Thus, our development process considered this unpredictability in our design scheme. First, as a team, we crafted a 15-minute long activity, creating utterances for Skusie to spark the children’s conversation and then imagining how children would respond. After observing triad interactions, the design team met, examined the video recording, and made adjustments to the scenario for the following sessions. In the example below, the children could not agree when asked to choose a birthday present for Skusie’s friend. The children repeated their own choices and were not able to reach an agreement by the end of the triad session.

ROBOT: Will you help me choose a birthday present for my friend? [both children lean forward to look at the picture]
BLJE: Un biciclo, un coche, unos juguetes- (A bike, a car, some toys-)
ROBOT: ¿BLJE- Cual debo darle a mi amigo? (BLJE, which should I give my friend?)
BLJE: Si es de tu tamaño, escoge un coche. (If it’s your size, choose the car.)
GWVI: You could get her a doll. The Barbie, with the dress-
BLJE: Que? (What?)
GWVI: With the dress.
As a result, the design team added statements to Skusie’s utterances in the following sessions to promote cooperation between children: “Can you two talk first and choose one for me?” and “Can you two choose together?” This simple invitational addition encouraged children to talk with each other to reach an agreement.

Our fourth design challenge was that children have short attention spans in general. It was very beneficial for the robot to call on children, ask questions, or repeat instructions. In the example A below, Skusie was able to get a child’s attention by calling his name and moving toward him. Having Skusie call on the child it wanted to invite into the conversation worked very well, especially for shy children and our bilingual pair. Skusie’s invitation led the children to take turns in their response. Also, Skusie expressed confusion and showed images on its smartphone brain to draw children’s attention. During the first part of the triad, GLAL, a Latina girl, was quiet and not overtly engaged. However, when Skusie showed pictures of her school on its smartphone brain, GLAL was immediately engaged as in the example B.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>BWTY: [Whispering to Moderator] How did he get here?</td>
<td>ROBOT: I saw lots of things on my way here. [The robot rolls forward and shows an image of the children’s school.]</td>
</tr>
<tr>
<td>Moderator: He came to visit me.</td>
<td>GLAL: That’s the gym!</td>
</tr>
<tr>
<td>BWTY: [Still whispering] How?</td>
<td>ROBOT: What is this place?</td>
</tr>
<tr>
<td>Moderator: From far, far away.</td>
<td>GLAL: A gym!</td>
</tr>
<tr>
<td>ROBOT: BWTY.</td>
<td>BWOL: That’s - that’s just like our gym!</td>
</tr>
<tr>
<td>GNSA: [Whispering] BWTY. Your turn.</td>
<td>ROBOT: Amazing. Thank you. Do you learn here?</td>
</tr>
<tr>
<td>ROBOT: Would you like to have pets?</td>
<td>GLAL and BWOL: Yes.</td>
</tr>
<tr>
<td>BWTY: Uh, yeah.</td>
<td></td>
</tr>
</tbody>
</table>

**Emerging themes in children’s interactions**

Our second question asked, *What themes arise in children's interactions with the robot and each other?* In examining the data, four main themes arose.

**Engagement with the robot**

When children first met the robot, they were generally curious and surprised. While some children were initially hesitant to talk with Skusie, as they met repeatedly with the robot and talked with each other about the triad experience, their engagement with Skusie grew. Often, they would lay on their bellies during the activity, with their heads near Skusie’s head. They asked the robot questions about previous scenarios and interactions, wanting to catch up with it after some time away. Also, most children were hesitant to leave the robot at the end of the activity. They tended to linger by the robot, asking questions about it or to it. The following examples illustrate this phenomenon.

**Treating the robot like a person**

Children talked to the robot directly, asked it questions, and took turns with it, modeling sophisticated, inclusive conversation skills. Often, they used its name or the second person pronoun “you.” Other times, they used the anthropomorphic pronouns “he” and “she.” BLED, for example, talked to Skusie directly, asking it, “Skusie, what do you know about tigers?” He then told it, “Sharks are very difficult, Skusie. They- they- they have big teeth.” Later, when his partner digressed from the conversation topic, BLED tried to get the conversation back on track by asking, “What’s going to happen next, Skusie?” BWHU asked Skusie for a high five. When co-constructing an interaction with Skusie, BWOL said, “Let’s take him on an adventure. Oh this is going to be the funniest day of school I’ve ever had.” He then dramatically fell down and popped back up to continue the activity.
Forgiving the robot’s weaknesses

Children were forgiving of Skusie’s imperfections. Their desire to help Skusie learn about life on earth seemed coupled with patience and support for its – and our – efforts. Children were generally patient and empathetic with Skusie even when it interrupted them and when it responded inappropriately due to software glitches or controller mistakes. Children were happy to meet with the robot, hugging and showing affection to Skusie.

GLAL: Skusie move!
BWOL: It’s ok. Skusie’s a robot. Skusie doesn’t even know about eating yet. ‘Cause he’s from a different planet, not Earth.

BWOL: What favorite animal do you like, Susie?
ROBOT: I don’t understand you.
Moderator: She doesn’t know yet. ‘Cause she’s still learning about all of our animals here on Earth.
BWOL: I’m still learning about other things.

BLED: “You already tried that Skusie.” [when Skusie repeated the same picture]

Learning to work together

When encouraged by the robot, children from the mainstream culture and from Latino families gradually learned to work together, sometimes across the language barrier. At the beginning of the triadic sessions, children habitually talked to the robot individually as if each were alone with the robot. Quite often, when asked questions by the robot, they answered simultaneously or gave Skusie opposite answers. In addition, shier children often lacked an opportunity to speak when they were paired with a more talkative child. After a few sessions where the team observed this phenomenon, we added some explicit statements to the robot’s utterances, such as “Can you talk one after the other?” “I am confused.” And “Can you two talk first and tell me one thing at a time?” These requests from Skusie usually induced children’s cooperation immediately. Such immediate improvement had not occurred in earlier human-brokering sessions. Toward the end of data gathering, in May, the robot’s communication skills had improved such that triads often had natural, easy conversations where all members - children and robot – contributed equitably.

Implications for designing for young children

A few implications are drawn for designing for children. First, explicit invitation, such as calling on and rolling over to the child, was helpful in gaining children’s attention overall and particularly in encouraging shy children to talk. Related to this, to elicit some children’s verbal responses, the robot repeated its invitational utterances. Quite often, children remained quiet and just gazed at the robot in response to its first utterances and spoke out only after two or three repetitions of the utterances. Also, children’s autonomous exploration should be encouraged in the interaction with the robot. Just as they voluntarily brought in their prior knowledge and experience to help the robot, children seemed to accept the robot’s free responses that were neither aligned with children’s own nor flowed logically. In this sense, the unpredictability of young children could be used in robot/child interaction design in a productive way. Next, a robot’s empathy was contagious. When the robot expressed appreciation for children’s information from their various personal experiences; children were also patient and understanding in their interactions with the robot and each other. Lastly, the bilingual robot seemed to help reduce stigma associated with language, perhaps, due to the study context uniquely. It was clear that when Skusie spoke to the children in Spanish, children from Spanish-speaking families participated more actively and responded to it gradually in Spanish.

References


Harré, R., & Moghaddam, F. (Eds.). (2003). *The self and others: Positioning individuals and groups in personal, political, and cultural contexts*.


**Acknowledgments**

This study was supported by the U.S. National Science Foundation (Cyberlearning and Future Learning Technologies #1623561).
Determinants of School Level Success in Design-Based Innovation Networks

Nancy Law, Yeung Lee, and Constance Wong-Loke  
nlaw@hku.hk, yeunglee@hku.hk, constancewongloke@gmail.com  
University of Hong Kong

Abstract: Design-based implementation research (DBIR) is a methodological approach that has emerged in recent years to address the challenge of sustainability and scalability of such learning innovations in diverse settings in the educational ecosystem. These efforts are generally organized in the form of innovation/implementation networks (DBIN). While the network as a whole can be successful, implementations in individual schools differ, resulting in large variations in the depth and reach of the changes that take place at the school level. This study investigates two schools in a DBIN that demonstrated similar trajectories of change in the first two years, but very different outcomes at the end of the third year. The findings show that the fine grain details of the architecture for and leadership engagement in within- and cross-school learning, and the quality of the school-based learning interactions matter in determining the implementation outcomes of schools in the same DBIN.

Introduction

Design-based implementation research (DBIR) (Fishman et al., 2013; Penuel & Fishman, 2012) has increasingly become a methodology of choice when the goal is to develop sustainable and scalable solutions to significant problems of practice through the disciplined application of research based theory and knowledge. DBIR needs to engage two layers of theory (Penuel, 2015). The first level theory is about how students learn and focuses on learning taking place at the classroom level, and design-based research (DBR) has been contributing to our understanding at this level. However, scalable implementation of pedagogical innovation needs to scaffold learning beyond the classroom to higher levels of the complex, hierarchically nested layers of the education ecosystem. An important underpinning principle of DBIR is the partnership relationship among the multiple stakeholders involved, usually constituted as an innovation network (DBIN). In this study, we explore the second level of theory from the perspective of school level scalability of innovation and change in a DBIN.

It has been shown that organizational and administrative support at the school level such as timetabling and space for co-planning (Gumus, 2013), opportunities for peer observation (DeLuca et al., 2015) are important. (Wenger, 1998) refers to the organizational environment that provide opportunities and supports for collaborative learning in a community of practice as the “architecture for learning”. Studies in educational change also found the architecture for learning, such as organizational routines, team structures and roles to be of critical importance to the effectiveness of educational change efforts in the classroom (Spillane, Parise, & Sherer, 2011; Stein & Coburn, 2008). The team structure, content, interaction foci, decision-making process and roles of participants are core elements of an architecture for learning (Law, Yuen, & Lee, 2015) in the context of teacher co-design.

This paper is an investigation into what differentiates implementation conditions and strategies between schools in the same DBIN that have successfully scaled from those that failed to do so. The study is underpinned by a model of scalable change as aligned learning at multiple levels, using the framework of architecture for learning. The context of this study is a DBIN that was funded as a university-school partnership program for three years. Two schools with apparently similar initial conditions and trajectories of implementation but widely different outcomes at the end of three years are selected as cases for in-depth analysis for this study.

The context of this study is a government funded 3-year university-school partnership program, titled “Self-directed learning in science with e-learning support for learner diversity and smooth primary-secondary transition” (SDLS), funded by the Education Bureau in Hong Kong. The overarching goal is to nurture and develop student’s self-directed learning ability to become confident and capable life-long learners in the 21st century. In this paper, we first introduce the nature and design principles of SDLS at the network (DBIN) level. This is followed by a description of the implementation trajectories in both schools over the three years. Finally, we present our analysis of the schools’ architectures for learning and school-level conditions of implementation.

Design principles for SDLS as a DBIN

As mentioned earlier, DBIR involves two layers of theories: theories about how students learn (classroom and individual level learning), and theories about organizations in the context of such changes. In the SDLS Project, self-directed learning (SDL) is the underpinning pedagogical theory guiding the design of learning at the student
and teacher levels. For SDL to be implemented in classrooms, it requires deep changes in teachers’ pedagogical practice, expertise in learning and assessment design, changes in routines, resource allocation, and other aspects in the architecture for learning (Wenger, 1998) for the pedagogical innovation to be sustained and scaled.

**Self-directed learning as the pedagogy of choice for the DBIN**

As a pedagogical approach, SDL has two key features that are characteristic of learning as collaborative inquiry: (a) it requires students to take personal ownership of their learning (setting learning goals, accepting responsibility for their thoughts and actions, and maintaining control for the many learning decisions), and (b) the learning process often involve interactions with the teacher as well as with other learners (Brockett & Hiemstra, 1991). Synthesizing the literature on SDL (Candy, 1991; Gibbons, 2003; Tan et al., 2011) and aligning SDL with the process of inquiry and problem solving, we have articulated a five-component model of SDL: goal setting, self-planning, self-monitoring, self-evaluation and revision, as the operational definition SDL for the DBIN. A brief description of these five components are presented in Table 1.

**Table 1: A description of the five SDL components**

<table>
<thead>
<tr>
<th>SDL component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal setting</td>
<td>Students identify own learning goals, targets</td>
</tr>
<tr>
<td>Self-planning</td>
<td>Students regulate and plan for the detailed decisions and arrangements associated with their own learning</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>Students self-manage their own time, monitor their own repertoire of learning strategies, and adjust their own learning pathways as they progress</td>
</tr>
<tr>
<td>Self-evaluation</td>
<td>Students are aware of the assessment criteria, and apply these for self-evaluation</td>
</tr>
<tr>
<td>Revision</td>
<td>Students draw on peer and/or teacher feedback to reflect on their own learning, and make appropriate revisions, and apply learning to new contexts</td>
</tr>
</tbody>
</table>

Critical to successful pedagogical change is deep teacher learning, learning that leads to deepening changes in teacher practice. In this DBIR project, SDL also underpins the learning design for teachers. Teachers need to experience learning as a self-directed learner in collaboration with other teachers: they need to be given the space to determine what they consider as appropriate and reasonable for their SDL experimentation in their own classrooms, plan, implement, monitor and reflect on the experience for further improvement and experimentation. We take a growth perspective on both student and teacher learning, and do not impose targets on teachers’ practice.

Teacher collaboration in teacher design teams provides opportunities for teachers to engage in “professional talk”: reflecting on the curriculum, their practices and goals for students’ learning (Parke & Coble, 1997); exploring new pedagogies, including the effective use of technology in teaching and learning (Agyei & Voogt, 2012). Opportunities for co-planning are not adequate to ensure productive teacher learning. The team membership composition in terms of background and expertise, team size (Handelzalts, 2009), the nature and focus of team interactions (Ehrleinspiel, Giapoulis & Günther, 1997), and the role of school-based facilitators in the design and associated decision making process (Huizingga et al., 2014) have been shown to have important implications for the learning outcomes of teacher design teams. In our project, we adopt teachers’ collaborative inquiry centering on the co-design of students’ learning, peer observations and reflective debriefing as the format for implementing SDL for teacher learning. The details of this design will be described in a later section.

**Social technical co-evolution and the multilevel multiscale model of aligned learning for scalable innovations**

Pedagogical change and innovation in a classroom does not happen in a vacuum. Classrooms are embedded in schools, communities and education systems and are influenced by intricate interdependencies with local, national, international communities and organizations as well as business enterprises (Davis, Eickelmann & Zaka, 2013). There is increasing recognition that the scalability of educational innovations can only be addressed if we can take account of the complexity of education ecosystems (Sabelli & Dede, 2013). As Law et al. (2016) illustrate (see Figure 1), different parts of the education ecosystem are highly connected and interdependent. For students to achieve the desired 21st century learning outcomes such as the 4Cs and digital literacy through collaborative inquiry and associated pedagogical and assessment innovations, aligned changes at teacher, school and system levels are needed. Such changes can be conceptualized as learning at these multiple levels. Hence, scalable innovation in DBIR is not a one step process, but an iterative process of aligned change through interactions and feedback across the different sectors and levels of the ecosystem. The second level of theory in DBIR is to address the challenge of achieving the necessary aligned learning.

Current literature on DBIR has generally taken on a sociocultural perspective in understanding the challenge associated with achieving alignment in the process of change, in particular, Lave & Wenger's (1991)
community of practice (CoP) theory. Members of the same professional practice share the same understanding of their enterprise, norms of mutual interaction and repertoire of capabilities. Scaling pedagogical innovations require communication across different communities of practice: teachers, principals, district level subject matter experts, policy makers, etc. To facilitate boundary crossing (Stein & Coburn, 2008) between different CoPs, there needs to be boundary practices (i.e. interactions constituted to bring members of different CoPs together on issues of common concern) and boundary objects (i.e. artefacts that scaffold boundary crossing). Stein & Coburn (2008) further put forward the concept of architecture for learning, which refers to the conditions conducive to learning across boundaries, and identified the following as key elements of the architecture: networks/organizations (formal and informal) that connect people within and across CoPs, mechanisms through which people interact, and the artefacts that serve as reifications that embodies a set of ideas or processes for sharing and communication within or across community boundaries.

While the architecture of learning provides the environmental conditions for learning, any design for learning needs to be underpinned by a theory of how learning happens. Based on the idea that deep learning within each CoP would most likely occur through inquiry learning with peers, Law, Yuen and Lee (2015) propose a multilevel multiscale (MLMS) model of learning, which argues that an architecture for learning for CoPs at higher levels of the education eco-system needs to span larger units of scale. The scale of a design-based innovation network that can support learning of a CoP of school leaders would have to engage multiple schools, and the learning of a CoP of district leaders would have to engage multiple districts. An implication of the MLMS model as a DBIR design principle is that the architecture for learning needs to scaffold learning within and across units at different levels of scale as well as across the different levels. In the SDLS Network, the MLMS model of learning serves as a design principle for the second layer of theory to guide the design-based implementation research.

Design of the architecture for learning for the SDLS DBIN
SDLS conceptualizes four levels of structure in its design: student, teacher, school and network, and the DBIR designs architectures for learning at the network level and encourages schools to design its own school-based learning architecture. At the network level, we provide a finer grain organizational structure to connect schools and teachers in multiple ways for different types of interaction foci and mechanisms to support scaling across the three years. The network started with 10 primary and 10 secondary schools, organized in two clusters of five primary schools and two clusters of secondary schools as illustrated in Figure 2. Assignment of schools to these clusters is based on their geographic proximity and are hence referred to as regional clusters. An additional five primary and five secondary schools were added to the network in each subsequent year, and the clusters were restructured so that there is a mix of old and new schools in each cluster.

There are four types of cross-school teacher-level learning activities organized for schools in the network:

1. Regional cluster meetings—these cluster meetings are held in one of the five schools within each regional cluster, and these generally focus on sharing around the day-to-day work of teachers in the design of curriculum units, learning and/or assessment tasks, choice and deployment of learning resources/tools (both conventional and digital). These meetings are generally informal and teachers find these to be very useful and enjoyable. Teachers in the same cluster often connect outside of the scheduled cluster meetings, usually through electronic means.
2. Primary school level and secondary school level cluster workshops involving all project schools at the same school level—these school level meetings are generally used to introduce curriculum and pedagogical ideas related to specific curriculum themes/topics of interest to the teachers in the form of inquiry tasks such that teachers will experience together with their peers in a playful manner the learning tasks as students. Teachers often find that they can pick up useful ideas and snippets for adoption in their teaching, and at the same time experience firsthand self-directed learning in science as a learner.

3. Whole network meetings (four in a year)—these meetings serve three key functions: (1) introducing self-directed learning as a pedagogical approach and its core features; (2) sharing of SDL learning designs and their implementation in classrooms by project school teachers, often including classroom videos, samples of students’ work, and teachers’ own reflections; and (3) introducing some “advanced topics” such as designing assessment as learning, using virtual reality tools for scientific exploration, etc.

4. Open classroom and peer observation of teaching—each school commits to designing and implementing one SDL curriculum unit per semester. University-based project staff are assigned to each of the schools and conduct school-based co-planning sessions for each implementation. Each school will select one lesson from one of the two implementations each year for open classroom observations by interested teachers and principals in the network. A debriefing session involving teachers in the observed school and the visitors will take place immediately following the observed lesson.

Cross-school teacher-level learning activities are scheduled throughout the school year so that each month one of the activities listed under 1 to 3 will be organized. The different meetings are interlaced so that teachers will have opportunities to engage in different collaborative learning activities that inform and/or support their own school-based collaborative inquiry of SDL implementation, culminating in the implementation of their own designed SDL units and the open classroom for peers’ observation. Principals of participating schools need to commit to releasing their teachers to participate in the monthly cross-school meetings and workshops.

Two of the four whole network meetings include the participation of principals. For these, half of the meeting is designed as opportunities for cross-school, cross-level interactions involving both teachers and school leadership (principal and/or vice principal) as it is important in the MLMS model that leadership gets opportunities to learn about how other schools implement SDL at classroom and school levels. During this half of the meetings, in addition to teacher sharing of their practices, principals and school project team members are also invited to share the management and resource allocation provisions and other strategic organizational decisions made at the school level to address challenges encountered and/or further ways to enhance the innovation implementation in their schools. This provides an opportunity for teachers and school leadership to understand the innovation implementation as a multilevel effort within a school. During the other half of these meetings, school principals will take part in “leadership circle” meetings separate from the teachers so that they can discuss and share challenges and solutions at the school level. Principals are also strongly encouraged to participate in at least one open classroom observation in another school in each year.

Methodology
In this study, we wish to understand how schools design their own architectures for learning (AfL) to support their implementation of SDLs, as well as whether and how features of the AfL relate to different implementation outcomes. The methodology adopted is comparative case studies of the network schools. We collect quantitative and qualitative data on (1) the schools’ participation in the various activities organized by the network; (2) the planning, implementation and post-lesson reflection for each SDL curriculum units designed and implemented in each semester by each school, through document collection and observations conducted during co-planning meetings, lesson observations, debriefing meetings, and online curriculum-related activities of the teachers and students on iLAP, a customized Moodle platform serving as the project Learning Management System; and (3) each school’s architecture for learning (both formal and informal) through surveys and interviews with the
principal and project core team members in each school at the beginning of their first year of participation and at the end of each project school year.

In this study, we have selected two primary schools (Schools A and B) that joined the project in Year 1 for our investigation. Both are government-funded schools located in public housing estates (government housing for low-income families). These two schools showed similar trajectories in SDLS implementation during the first two years in the project but differed greatly in scale and depth of implementation by the end of the third year.

**The trajectory of SDLS development in the schools**

In Year 1 of the SDLS project, participating primary schools were advised to start their innovation implementation in Grade 5 so that the teacher workshops and cross school teacher collaboration can be focused around topics in the Grade 5 science curriculum. We required all schools to involve at least two teachers at the same grade level in the implementation of SDLS so that there can be peer collaboration within schools. Schools were encouraged to engage all GS teachers at the implementation grade levels for sustainable school-based adoption of SDLS.

Schools A and B decided in Year 1 to involve all the grade 5 classes in their schools (A has three classes at each grade level and B has four). Both also extended the implementation to include all grades 5 and 6 classes in Year 2. By the end of Year 2, there were discernible shifts in teachers’ pedagogical and assessment practices in both schools from teacher-directed practices to more student-centered practices. The teachers were willing to let go of control and let students take the lead in designing experiments from the beginning to the end. There was also a shift from teacher-based assessment to include peer assessment. In School B, students were asked to design their own self-assessment rubric. The teachers also reported their own learning gains:

“I have lots of questions in making the electric fan. … as I am not a [science] subject specific trained teacher, in the process of creating a powerful fan, I engaged as a self-directed learner in solving the problem and learnt some scientific concepts. My scientific knowledge was enhanced. I am more confident as I went through what my students will encountered in the learning processes.” (School A teacher)

“Traditionally, we design the worksheet for students to fill in the blanks. Now students need to design their own experiments, they need to set their goals, steps and review their results. Sometimes, they encountered problems and failed. It is good that students have this kinds of experiences ….” (School B teacher)

 Apparently, these two schools had similar implementation scales and the teachers had similarly found the SDL learning experience to be beneficial both to their students and themselves. However, the two schools diverged in their implementation trajectory in the third year. School A kept the same project core team, maintained its momentum and expanded the implementation to include all classes in grades 4, 5, and 6. The school further created more e-learning courserooms on the iLAP beyond the ones set up by default for each participating class. Some of these additional courserooms were for students’ SDL explorations on topics outside of the school curriculum, and some for teachers who wish to try implementing SDL with e-learning in non-science related subjects. Teachers in School A continued to deepen their understanding of SDL and gave progressively more agency to their students. The teachers observed that when students were interested and given choices to set their own goals, they were motivated to try different ways to accomplish those goals. At the end of Year 3 interview, one teacher reported becoming more self-directed, and that she had learned more from her students, particularly when they asked unexpected questions. Several of the School A teachers felt that the self-directed learning accomplished by students in all classes and ability groups exceeded their expectations, even though at the beginning of the year they faced the challenge of having students whose cognitive and creative abilities varied greatly.

School B on the other hand decided to change the personnel involved in the project, including the team leader and the scale of the implementation. The original project leader was moved to be in charge of another project, and only 2 classes at Grade 4 were involved in SDLS in this final year. The rationale given for the change in scale was the lack of adequate ICT infrastructure in the school. The SDL practices implemented in the two P.4 classes were also more teacher-centered compared to the previous two years. There were essentially no e-learning activities included in the SDL implementation. The two teachers involved did not have experience with iLAP and they did not join the training workshops offered for teachers new to the project. At the end of year 3, these teachers found students’ SDL learning outcomes to be mediocre, and they attributed this to the weaker capacity of P.4 students to undertake SDL.

**Architecture for learning at the school level**
Why did the two schools’ implementation trajectories diverge so drastically in the third year? There was no change in the leadership or staffing at the school level in either A or B. What may be the reasons for the sudden change?

In this section, we describe the AIL associated with the implementation of SDLS at each school over the three years. We first describe the project implementation team structure, interaction mechanisms, and decisions made by the two teams. This is followed by a description of the school-based teacher learning organization and mechanisms, and how these connect to the learning opportunities provided by the architecture for learning at the SDLS network level. Finally, we report on the leadership (principal and senior management team) engagement in project-related learning at the school and network levels.

**School level project team structure, roles, interaction mechanisms and decisions**

*Project team structure.* The School A project implementation team was led by the PSMCD (a formal, school-based curriculum leader role in primary schools) throughout the three years. The School B team was also led by its PSMCD for the first two years, but leadership shifted to the assistant principal in the third year. The PSMCD in both schools had curriculum and pedagogical expertise but did not have a science disciplinary background. The School B assistant principal had a science background. Other members of the School A project team included both the Principal and Vice principal, the General Studies (GS, the subject home for science in the primary curriculum in Hong Kong) panel head and all of the teachers involved in the project implementation during each of the years. Other School B project team members included the GS panel head and all of the project implementation teachers in each of the years. Hence, structurally, school A had top level leadership involvement, whereas School B only had involvement up to the middle management level.

*Project team participation in activities organized by the network.* School A makes provisions for as many teachers as possible to participate in the various cluster and whole network activities organized by the project throughout the three years. For School B, there was also active participation in the activities organized by the SDLS network during the first two years by project team members, except the principal. In year 3 only one or two teachers participated in the various project activities.

*Project team leader roles in learning.* The roles of the project team leaders differed right from the first year, even though the leader was the PSMCD in both schools. Both team leaders played the roles of liaison with the university-based network team and coordination of within-school project administration. In addition, the PSMCD in A served as the pedagogical mentor in the project and participated in all the school-based co-planning meetings as well as in all of the activities organized by the SDLS network, helped to resolve difficulties encountered in the implementation, including communicating issues and needs to the principal and the senior management team. In the third year, the PSMCD was so convinced of SDL as a pedagogy of choice for student learning in general that she started implementing SDL in her own Chinese language classes using iLAP. On the other hand, the role played by the PSMCD in B was only administrative, which included also the allocation of teachers to the implementation classes and arranging tablets for students’ use during classes as needed. In the third year, the new project leader (assistant principal) rarely participated in any of the activities organized for the clusters or the entire network. Hence, the project leader in A engaged fully as a learner and a leader of learning for the team, while the role of the project leader in B was solely administrative.

*Project team decisions.* As the PSMCD in A was fully involved in all of the learning activities at both network and school levels, the whole project team was involved in discussing issues, and making or facilitating decisions that would support project implementation and teacher learning. These included getting the school’s agreement to remove one curriculum topic to allow more time for student exploration, making arrangements for teachers to attend within-school and cross-school open classroom observations, creating additional course rooms in iLAP for more SDL focused e-learning activities for students, and creating a WhatsApp group to facilitate efficient communication among team members. In B, the project team did not make any specific decisions or recommendations to the senior management. The team discussed administrative arrangements to implement the participation requirements required by the university-led SDLS project.

**School-based teacher learning organization and mechanisms**

Both schools had grade-level teams involved in the co-design of SDL lessons in selected curriculum units. In A, all involved teachers contribute materials to the co-design work. At the co-planning meeting, which was held during the weekly timetabled co-planning period, the teachers discussed with the PSMCD and the GS panel head their lesson designs in detail, including the feasibility of the plan, anticipated problems and solutions. Experiments were tried out many times with different variables. Besides serving as the pedagogical mentor, the PSMCD also communicated to the principal for discussion at the senior management level if there are obstacles encountered in the implementation process. Teachers felt safe and supported during the implementation process, describing the PSMCD as a ‘pillar’ of support. Frequent informal interactions were facilitated by staff room seating arrangements.
and the use of WhatsApp groups. In addition to the formal open classroom observations, all project teachers took
turns to observe each another, provide peer feedback and made iterative revisions to their co-designed lessons.

There is no timetabled co-planning period in School B, where co-planning took a more product-oriented
approach. One teacher took the key responsibility of designing the SDL lessons, and sought the inputs of other
GS teachers at the same grade level to refine the plan during co-planning. In Year 2, the co-planning took on a
stronger learning focus, when GS panel head was also teaching Grade 6. There was a lot of informal discussions
and collaboration between him and the teacher with the key responsibility for SDL lesson design. These intensive
collaborative design interactions brought visible changes in the quality of the designed lessons and the scope for
student agency in the planned learning activities. It is unfortunate that these quality learning opportunities initiated
autonomously by these two teachers and the ensuing outcomes were unnoticed by the school leadership, and
terminated when the school changed the implementation staffing and scale.

Leadership engagement in SDLS-related learning at the school and network levels
While a school is not led by a single person, the school principal plays a very important role in setting the direction,
communicating and implementing the school’s priorities, serving as a role model and a cheer leader. The principal
in School A was almost a classic example of a fully supportive school head. He attended all the open class
observations and debriefing sessions held in his own school, participated in some of the co-planning meetings in
his own school, offered his school as the venue for regional cluster meetings and occasionally attended these,
observed some open classrooms in other project schools every year, attended every year the two network level
meetings to which principals were invited, and shared his school’s project implementation experience with other
schools during whole network meetings. The principal in School B attended all the three open classroom
observations in his own school, and only participated in one whole network meeting in Year 1.

Innovation implementation often require changes in school routines (Spillane et al., 2011), resource
allocation, as well as curriculum and assessment policy decisions. In A, there are examples of different categories
of strategic arrangement in support of SDLS implementation in the school. In terms of resource allocation, each
participating teacher was given a one lesson reduction in teaching load per week to recognize the extra time they
need to spend on the project. Technical support staff was allocated to the project to provide e-learning design and
on-site classroom support. The physical seating arrangement in the staff room was changed to allow the project
teachers to sit in close proximity so that they can easily have frequent informal interactions. Weekly timetabled
co-planning periods were instituted as a school routine to ensure that teachers not only can, but also have the
accountability to engage in collaborative co-planning. School-based peer lesson observations was also instituted
as a school routine to scale up teacher learning and to showcase good practices within the school. The school has
also modified its school-based assessment policy so that 10% of subject scores are allocated for project work
performance. It is clear that there is a conscious policy in A to support teacher learning within and across schools.

In school B, resource allocations similar to those in A were made: a one lesson per week reduction in
teaching load for participating teachers, and assignment of technical support for e-learning implementation.
However, no other provisions were made for the project implementation beyond making arrangements for teachers
to attend the minimal number of cross-school open classroom observations stipulated by the project.

School-level support for e-learning in SDLS
The iLAP online portal was designed to facilitate student learning, collaboration, self and peer assessment, as well
as making visible the students’ learning process and outcomes to themselves, their peers, teachers and parents.
Both schools made use of iLAP in the first two years, with School A making more sophisticated uses to facilitate
student interactions. In year 3, e-learning use in A escalated, but stalled in B. The school-based provisions for e-
learning also differed. School A invested in more class sets of tablets, ensured that teachers were trained to use
iLAP and the tablets, as well as provided training to students on Chinese character input to ensure that they could
use their home computers to work on the e-learning tasks on iLAP. In School B, the lack of ICT infrastructure
was given a reason to scale down SDLS implementation in Year 3, and students’ inability to do Chinese character
input on computers was mentioned by teachers as a hurdle to implement e-learning in the school.

Discussions
In this paper, we have analyzed the differences in the architecture for learning between the two schools in their
implementation of SDLS. The DBIN level support were the same for both schools. However, school A had a rich
and multi-thronged organizational and leadership structure as well as interaction mechanisms and environment
that allow them to take full advantage of the learning opportunities made available at the network level, which at
the same time fostered a strong collaboration and peer learning culture within the school. School B also provided
the same tangible resources to support the innovation, but did not demonstrate an awareness of the importance of
supporting within and cross-school learning in the implementation process. In changing the project leader and staff in the third year, the school has in fact undermined the human and social capital (Hargreaves & Fullan, 2012) that had begun to build up through the project. It is taken for granted that the design principles underpinning the student level learning design need to be made explicit to teachers so that they can be intentional collaboration partners in DBR. We propose here that in DBIR, there is a need for us to also make explicit the second layer of design principles for our collaboration partners, which should also be a focus for the university-school partnership.

References


Acknowledgements

Nancy Law acknowledges that her contribution to this work is supported by the Hong Kong Research Grants Committee Humanities and Social Sciences Prestigious Fellowship Scheme # 37600414.
Informing the Design of Teacher Awareness Tools Through Causal Alignment Analysis

Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven
kjholste@cs.cmu.edu, bmclaren@cs.cmu.edu, aleven@cs.cmu.edu
Carnegie Mellon University

Abstract: Designers of teacher awareness tools (e.g., dashboards) must not only anticipate the effects analytics will have on awareness, but also how this enhanced awareness might affect teacher decision-making, and in turn, student learning. Yet teacher awareness tools are not commonly optimized to guide teacher behavior in ways that are productive for learning. In this paper, we introduce Causal Alignment Analysis (CAA), a framework for data-informed, iterative design of teacher awareness tools, which links the design of awareness tools to educational goals. We illustrate the usefulness of CAA with a case study, demonstrating the successful design alignment of an awareness tool with a causal path from teacher tool use to student learning. Over a sequence of four pilot studies conducted in both simulated and live classrooms, we demonstrate the iterative refinement of Lumilo, a real-time awareness tool, to draw teachers’ attention towards students who may benefit most from a teacher’s help.

Introduction

Supporting teachers in orchestrating complex classroom activities has been identified as a key research and design challenge for the learning sciences community (STELLAR, 2011; Tissenbaum et al., 2016). In recent years, several real-time awareness tools have been designed and developed to aid teachers in orchestrating complex technology-enhanced learning scenarios (e.g., Alavi & Dillenbourg, 2012; Holstein, Hong, Tegene, McLaren, & Aleven, 2018; Martinez-Maldonado, Clayphan, Yacef, & Kay, 2015; Mavrikis, Gutierrez-Santos, & Poulouvassilis, 2016). These tools augment teachers’ “state awareness” during ongoing learning activities (Rodríguez-Triana et al., 2017). For example, such tools may present teachers with real-time analytics on student knowledge, progress, and metacognition within educational software (Tissenbaum et al., 2016).

The design and development of real-time teacher awareness tools is often motivated by an assumption that enhanced teacher awareness will lead to improved teaching, and ultimately, to improved student outcomes. Yet there is a paucity of evidence to support these claims, and scientific knowledge about the effects that such tools may have on teaching and learning in real educational settings is scarce (Molenaar & Knoop-van Campen, 2017; Rodríguez-Triana et al., 2017). As such, it is a challenging problem to design effective teacher awareness tools. Designers must not only anticipate the effects analytics may have on teacher awareness, but also how this enhanced awareness might affect teacher behavior, and how these changes in behavior will ultimately influence student learning. Compounding these challenges, while existing design workflows such as LATUX (Martinez-Maldonado, Pardo, Mirriahi, Yacef, Kay, & Clayphan, 2016) support the user-centered design of awareness tools based on teacher feedback, there is a lack of standard methodology for the outcome-driven improvement of awareness tools, to achieve targeted educational goals. Furthermore, justifications for design decisions (e.g., what information to present in a dashboard) are rarely reported in the literature (Rodríguez-Triana et al., 2017).

Researchers in other areas of educational technology research have adopted data-informed approaches to iteratively guide the design of technologies towards educational goals (e.g., Koedinger, Stamper, McLaughlin, & Nixon, 2013). For example, the design of intelligent tutoring systems (ITs) sometimes includes an iterative refinement process, in which historical student data is leveraged to increase alignment between the software’s instructional design and the way students actually learn the material, as inferred from data (e.g., Liu & Koedinger, 2017). By contrast, while teacher awareness tools are sometimes designed to be useful and usable, they are not typically optimized to guide teacher behavior in ways that are productive for learning. Given the complexity of designing teacher awareness tools, and the substantial causal distance between enhancing teacher awareness and enhancing student learning (Xhakaj, Aleven, & McLaren, 2017), bringing such outcome-driven approaches to the design of teacher awareness tools may be key to ensuring their effectiveness. In this paper, we introduce Causal Alignment Analysis (CAA): a framework for the data-informed iterative design of teacher awareness tools. We illustrate CAA via a case study, demonstrating the iterative improvement of a real-time awareness tool over a sequence of pilot studies. Finally, we discuss conclusions and directions for future work.

Causal Alignment Analysis for teacher awareness tools
Beginning from a specification of educational goals (e.g., improving student learning or engagement), CAA involves gradually aligning the design of a teacher awareness tool with these goals, by repeatedly evaluating the tool’s effects along hypothesized causal paths from teacher tool use to targeted student-level outcomes. Specifically, CAA begins by generating answers to the questions below, which may represent open hypotheses where theory is absent or underspecified:

1. What student outcomes do we wish the teacher awareness tool to support?
2. What student-level processes promote or hinder progress toward the goals specified in (1)?
3. What teacher-level processes promote or hinder the student-level processes identified in (2)?
4. How can an awareness tool better support and promote the processes identified in (2) and (3)?

Taken together, answers to these questions specify hypothesized causal paths from a teacher’s use of a particular awareness tool to enhanced student outcomes (as in Figure 1). Making the goals and hypothesized mechanisms of action of an awareness tool explicit early on may usefully constrain the design of an initial prototype. Once an initial prototype has been developed, Causal Alignment Analysis then involves prototyping the tool with teachers and students. Using data from these prototyping sessions, designers evaluate the alignment (or lack thereof) between the prototype’s observed effects on teacher behavior, and one or more hypothesized causal paths to improved student outcomes (cf. Xhakaj et al., 2017). Based on this analysis, designers can then refine the prototype with the goal of increasing alignment, thus increasing the chances that the tool will have a positive impact in the classroom. Finally, the prototyping cycle repeats, to evaluate the effectiveness of this realignment.

Figure 1 shows examples of potential causal paths from a teacher’s use of an awareness tool to teacher and student level outcomes. For these examples, we consider the context of self-paced classrooms in which students work with educational software, while a teacher uses a real-time awareness tool to decide when, with which students, and how to provide additional assistance. From left to right, the diagram shows potential influences of a teacher awareness tool (Q4) on the behavior of the teacher using it (Q3), potential impacts of resulting shifts in teacher behavior on students (Q2), and finally, potential impacts of these student-level effects on student learning outcomes (Q1). Given that a teacher has limited time to provide one-on-one assistance, the top path in Figure 1 posits that if teachers were alerted to critical situations (e.g., a student exhibiting a common misconception), they would be able to more effectively allocate time to students who need their attention the most, at the right moments (see Martinez-Maldonado et al., 2015). Thus, an awareness tool should be designed to alert teachers of such critical situations. In contrast to the top path — which represents a hypothesis that students using educational software would learn more from additional teacher assistance in certain situations — the second path, represents the hypothesis that students would benefit from more teacher attention, in general. Under this hypothesis, an awareness tool should be designed to encourage teachers to spend more time working with students, overall — perhaps by making teachers feel more informed, and thus increasing their overall “confidence to act” (van Leeuwen, Janssen, Erkens, & Brekelmans, 2015). The third causal path represents the hypothesis that, if the quality of a teacher’s one-one-one interactions with students were improved (e.g., more tailored to a student’s specific weaknesses), this would enhance student learning with the software (see van de Pol & Elbers, 2013). Furthermore, this path posits that if teachers were made more aware of student difficulties, this would lead teachers...
to tailor their one-on-one interactions more closely to individual students’ needs. The fourth causal path posits a direct link from a teacher’s use of an awareness tool and a student-level effect. Under this hypothesis, students’ mere awareness that a teacher is monitoring their activities in the software contributes to their learning, perhaps by increasing engagement (Holstein, McLaren, & Aleven, 2017b). Finally, the bottom path represents a hypothesis that teachers’ use of a particular awareness tool positively impacts their classroom experience (Rodriguez-Triana et al., 2017), but has no notable effects on student outcomes.

Despite showing a relatively small set of hypothesized paths – each specified at a high level of abstraction – Figure 1 illustrates the enormous breadth of the design space for teacher awareness tools. Focusing on different combinations of these paths may yield radically different tool designs. In addition to guiding the initial design of a teacher awareness tool alongside user-centered design methods (Holstein et al., 2018; Martinez-Maldonado et al., 2016), CAA can be used to inform the refinement of an existing awareness tool. A designer applying CAA to the refinement of an existing awareness tool would begin by considering the tool’s educational goals, and then work backwards from these goals (cf. Wiggins et al., 2001) to construct one or more hypothesized causal paths originating from a teacher’s use of an awareness tool (guided by existing data and theory where possible). By prototyping the awareness tool, and collecting outcome data, the designer would evaluate whether the tool is likely to have desirable effects along each node in the path, adjusting the design as needed. To illustrate the use of CAA in practice, we next demonstrate the iterative design improvement of a real-time awareness tool.

**Background: Co-design of a real-time teacher awareness tool**

In our prior work, we designed a real-time awareness tool for teachers working in K-12 classrooms using intelligent tutoring systems (ITSs): a class of advanced learning technologies that provide students with step-by-step guidance during complex problem-solving practice. ITSs have been found, in several meta-reviews, to enhance student learning in classroom settings, compared with other learning technologies or traditional classroom instruction (e.g., Kulik & Fletcher, 2016). A key benefit of using ITSs in the classroom is that they free teachers to circulate throughout the room, providing more individualized help while students work with the software at their own pace (Schofield, Eurich-Fulcer, & Britt, 1994). However, ITSs are not typically designed to support teachers in helping their students (Holstein, McLaren, & Aleven, 2017a).

We decided to focus our awareness tool design largely on the problem of supporting teachers in allocating scarce time and attention to those students who need it the most (the top path in Figure 1), during classes in which students work individually with ITSs. This focus was motivated, in part, by user-centered design work with middle school math teachers, which highlighted these decisions as a major challenge in orchestrating personalized learning (Holstein et al., 2017a; Martinez-Maldonado, et al., 2015). In addition, this focus was motivated by prior empirical results, suggesting that teachers’ decisions about whom to help, and when, may be impactful (e.g., Martinez-Maldonado et al., 2015). In particular, we focused on designing an awareness tool for classrooms using Lynnette, an ITS for equation solving (Long & Aleven, 2017).

In the first phase of our design process, we wanted to better understand teachers’ expressed needs and desires for real-time analytics. We adopted a participatory design approach, working closely with 16 middle school math teachers (across 9 schools and 6 school districts, in a large U.S. city and surrounding areas). We directly involved teachers at each stage of the design process (cf. Martinez-Maldonado et al., 2016), including the selection and tuning of analytics through iterative user testing (Holstein et al., 2017a, 2018). The initial prototype that emerged from this iterative process was a pair of mixed-reality smart glasses (Figure 2, top-right) called *Lumilo*, which displays real-time indicators of students’ current learning, metacognitive, or behavioral “processes” (as shown in Figure 2, left), floating above students’ heads (Figure 2, bottom-right), while allowing teachers to keep their heads up and attention focused on the classroom (Holstein et al., 2018). The indicators displayed by the initial prototype of *Lumilo* were ideas generated and iteratively refined by teachers, and implemented using established student modeling methods (e.g., Beck & Gong, 2013; Desmarais & Baker, 2012). Together, these indicators can be taken to represent, in part, the phenomena that teachers expect require their attention and/or intervention. For example, four teachers argued that alerts about high local error would require immediate intervention. Otherwise, these teachers worried that repeated error-making in an ITS might entrench the errors, despite negative feedback from the software (see Metcalfe, 2017). Teachers also found some indicators valuable for other reasons. For example, we found that positive indicators about student performance were valuable to teachers, in part, because they found them personally motivating (Holstein et al., 2018).

**Iterative improvement of Lumilo, using Causal Alignment Analysis**

In addition to serving teachers’ expressed needs and desires, however, we want to design awareness tools that can measurably benefit students. Teachers’ intuitions about the most important opportunities for intervention may not always be correct (e.g., Baker, Walonoski, Heffernan, Roll, Corbett, & Koedinger, 2008). Therefore, in the next
phase of our design process, we used Causal Alignment Analysis to iteratively refine Lumilo’s design, to increase its chances of having a positive impact in the classroom. With respect to the first of CAA’s four guiding questions, we had defined our learning objectives as the set of equation-solving skills that Lynnette tutors. In answer to CAA’s second and third questions, we adopted a causal model search approach to understand the relationships between Lumilo’s indicators and student learning outcomes – hypothesizing that teacher attention should be directed to student processes with a negative influence on learning. Finally, in response to CAA’s fourth question, we iteratively refined Lumilo to direct teachers’ time and attention towards these processes, over a sequence of in-lab and classroom pilot studies. Each step is discussed next, in turn.

To answer CAA’s second question (“What student processes promote or hinder ...”), we sought to better understand the relationships between student processes detected by the current prototype of Lumilo (the student-level indicators shown in Figure 2, emerging from our participatory design process) and student learning within Lynnette. To this end, we adopted a causal model search approach, using directed acyclic graphs (DAGs) to represent the causal structure among variables measured by Lumilo, and student assessment scores. We collected data from 115 middle school math students (across 7 classrooms and 4 teachers), each of whom worked with Lynnette for 60 minutes. In these classrooms, the teacher did not use an awareness tool (Table 1, Study 1). In all studies, we assessed students’ equation-solving skill with a pretest and posttest administered before and after using the tutor. We used two forms that were identical except for the specific numbers used in equations. We presented the forms in counterbalanced order across pre- and posttest.

We then used the PC algorithm in the Tetrad V program to search for an equivalence class of DAGs, consistent with a set of conditional independence constraints (Spirtes et al., 2000). The PC algorithm is asymptotically reliable; its primary limitations are its assumptions that no unmeasured confounders are present, and that any underlying causal relationships between variables can be modeled by linear functions. To relax the former of these assumptions, we also used the FCI algorithm, which allows for the possibility of unmeasured confounders. The FCI algorithm learns an equivalence class, represented by partial ancestral graphs (PAGs), encoding uncertainty over the nature of pairwise relationships between variables (Spirtes et al., 2000). To inform both searches, we provided background knowledge about our study design as a search constraint: we specified that the pretest was prior to any process variables, and that all process variables preceded the posttest.

Figure 3 (left) shows the DAG learned with the PC algorithm, including normalized coefficient estimates, to enable comparison of magnitudes. This model suggests that, of the indicators included in the initial prototype of Lumilo, three are potential direct causes of reduced student learning within the software: help abuse or gaming-the-system (measured by the Help Model and gaming detector, reviewed in Desmarais et al., 2012), high local error (defined by teachers as an error rate greater than 80%, within the last 8 student actions on the current activity), and unproductive persistence (measured by the “wheel-spinning” detector, described in Beck & Gong, 2013). This model fits the data well ($\chi^2 = 18.33, df = 19, p = .50$) (1). Figure 3 (right) shows the PAG learned with the FCI algorithm. In this figure, bidirectional links indicate the presence (and circle-origin links indicate the possibility) of unmeasured confounders. Otherwise, links indicate causal relationships. Wide links indicate no unmeasured confounders, and dark, wide links further indicate direct relationships. The PAG equivalence class found by FCI suggests that unmeasured confounders could potentially explain several of the links between...
Lumilo’s indicators. Finally, in both causal models, gaming/help-abuse, unproductive persistence, and help avoidance (as measured by the Help Model, Aleven et al., 2016) are negatively linked to student learning. The model found by FCI suggests that out of 7 negative teacher-generated indicator ideas implemented in Lumilo, only one is directly linked to student learning: unproductive persistence. Influences of help avoidance and gaming/help-abuse on learning may in turn be mediated through unproductive persistence.

To determine how the design of Lumilo might be improved (the fourth question in CAA), we wanted to first understand how the current prototype of Lumilo influences teacher behavior, prior to deploying it in real classrooms. To this end, we conducted a series of simulated class sessions using a new prototyping method called Replay Enactments (REs) (Holstein et al., 2018). In each session, historical student interaction data were replayed in ITS interfaces, on separate computer screens in a classroom setting (but with no actual students present). Following a 35-minute training period in which teachers acclimated to using the tool and studied the definitions of each of Lumilo’s indicators, each teacher participated in a 40-minute replay session. In these sessions, teachers wore Lumilo, and were asked to think aloud while monitoring the “class”. If a teacher thought they would intervene with a certain “student” at a given time, the teacher would approach that “student” and enact the help session aloud. In addition to recording think-aloud data, we used Lumilo to automatically track the teacher’s physical position moment-by-moment (Holstein et al., 2018).

First, we investigated how teachers’ time allocation across students during REs may have been influenced by each of Lumilo’s student-level indicators. Teacher time allocation was measured per student by the cumulative time (in seconds) spent within a 4-ft. radius of that student (resolving ties among students by proximity), as well as time spent monitoring the student’s activities via Lumilo’s deep-dive screens (Holstein et al., 2018). Table 1 (Study 2) shows group-normalized correlations between detected student processes and teachers’ time allocation during six REs. Real-time indicators that were not significant predictors of teacher time allocation are omitted. As shown, occurrences of four of Lumilo’s indicator alerts were significantly positively correlated with teacher time allocation. Second, to understand the degree to which the awareness tool might have directed teachers towards students most in need of help, as per the top path in Figure 1, correlations between student assessment scores and teacher time allocation are also shown in Table 1. Given that teachers did not have access to assessment scores during REs, and that it is not possible to influence learning during a replayed class, we take the correlation between teacher time allocation during REs and student posttest scores (controlling for pretest) as evidence that Lumilo can direct teachers’ time to students who would otherwise exhibit lower learning. However, this correlation was relatively small, suggesting room for improvement.

Taken together, these analyses suggested various ways the design of Lumilo could be improved (Q4), to increase its alignment with the hypothesized causal path shown in Figure 4. Unproductive persistence was the weakest driver of teacher attention during REs, out of the indicators correlated with teacher time allocation (as shown in Study 2 of Table 1), despite being the one variable directly (and negatively) related to student learning in the causal model found by FCI. To better align Lumilo’s design with these analyses, the design should focus more explicitly on alerting teachers to cases of unproductive persistence, by increasing the salience of this alert and others that may serve as reliable early predictors. For instance, although help avoidance is a potential cause of unproductive persistence in the PAG found by FCI (and thus potentially valuable as an early predictor), it was
not a significant driver of teacher attention. Similarly, this model suggests that less emphasis should be placed on alerting teachers to high local error or rapid attempts in general, and more should be placed on alerting teachers to cases that constitute maladaptive help-use and/or gaming (Desmarais & Baker, 2012). As such, we next refined the prototype of *Lumilo* to place greater emphasis on alerts about unproductive persistence and persistent help avoidance. This included not only making the corresponding indicator symbols more visually salient than others (larger and brighter), but also drawing teachers’ attention to these alerts through ambient sound notifications. Meanwhile, we de-emphasized other alerts, including high local error and rapid attempts by making these indicators relatively dimmer and smaller. Furthermore, if a student was detected as unproductively persisting on one or more skills, avoiding help, or gaming/abusing-help, any other alerts for that student would be hidden. We ran additional classroom pilots using *Lumilo* (version 3) in 4 classrooms. Students in each classroom worked with *Lynnette* for 40 minutes, while the teacher used *Lumilo* (version 2) to monitor and help students. Students’ domain knowledge in equation solving was measured before and after using the software, via computer-based pre- and posttests, as in prior studies. As shown in Study 3 of Table 1, students who were more frequently detected as unproductively persisting or avoiding help received significantly more teacher time during this single-classroom pilot, compared with students exhibiting other behaviors tracked by *Lumilo*, suggesting that the design refinements may have had the intended effect. Furthermore, the teacher’s attention during this single-classroom pilot was strongly and significantly focused towards students with lower prior domain knowledge (as measured by the pretest), and the correlation between teacher time allocation and student posttest score (controlling for pretest) was positive, despite a likely selection effect, although not statistically significant.

Following this pilot, we made minimal design refinements to *Lumilo*, in an effort to ensure that alerts of unproductive persistence were emphasized (as potentially more critical) over alerts of help avoidance and gaming/help-abuse. In version 3 of *Lumilo*, if a student was detected as unproductively persisting in the software on one or more skills, any other alerts for that student would be hidden. We ran additional classroom pilots using *Lumilo* (version 3) in 4 classrooms. Students in each classroom worked with *Lynnette* for a total of 60 minutes while the teacher used *Lumilo* to monitor and help their students. As before, student domain knowledge was measured via 20-minute, computer-based pre- and posttests. As shown in Study 4 of Table 1, unproductive persistence was the strongest predictor of teacher time allocation, followed by help avoidance and gaming/help-abuse. Classroom observations indicate that teachers continued to make use of all indicators presented by *Lumilo* (e.g., praising recent high performers or nudging inactive students), but tended to reserve in-depth remediation sessions for those students detected as unproductively persisting. Retrospective post-interviews corroborated this observation. However, teachers also reported frequently attending to “quick fix” alerts for students physically “en-route” to a particular student the teacher was targeting for remediation.

In summary, in the first phase of our design process, we decided to focus on the problem of supporting teachers in allocating scarce time and attention to those students who may need it most. We adopted a participatory design approach, eliciting ideas for real-time analytics that teachers considered actionable, relevant to learning, or otherwise valuable to monitor. We leveraged pre-existing student modeling techniques to provide teachers with

---

**Table 1: Correlations between teacher time allocation, and detected student processes and test scores, \(^{*}p < 0.05, **p < 0.01, ***p < 0.001.**** Rows show a series of studies, using successive versions of *Lumilo***

<table>
<thead>
<tr>
<th>Study</th>
<th>Process Variables (awareness tool alerts)</th>
<th>Assessment Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study Context</td>
<td>Unproductive persistence</td>
</tr>
<tr>
<td>1</td>
<td>live none</td>
<td>(4, 7, 115)</td>
</tr>
<tr>
<td>2</td>
<td>RE Lumilo v1</td>
<td>(6, 3, 90)</td>
</tr>
<tr>
<td>3</td>
<td>live Lumilo v2</td>
<td>(1, 1, 15)</td>
</tr>
<tr>
<td>4</td>
<td>live Lumilo v3</td>
<td>(2, 4, 84)</td>
</tr>
</tbody>
</table>
these analytics, while iteratively prototyping them with teachers to ensure their usefulness and usability. In the next phase of our design process, we used CAA to iteratively align Lumilo’s design with a hypothesized causal path to improved learning outcomes, learned from data (a finer-grained instantiation of the top path in Figure 1, as shown in Figure 4). With respect to the first of CAA’s four guiding questions, we defined students’ learning objectives as the skills that Lynnette is intended to tutor, and assessed student learning with respect to these skills. In answer to CAA’s second and third questions, we adopted a causal model search approach to discover a critical subset of Lumilo’s indicators, representing student processes that most strongly influence learning outcomes with Lynnette. In turn, we hypothesized that students exhibiting these processes may benefit most from out-of-software, teacher interventions. Finally, with respect to CAA’s fourth question, we iteratively refined Lumilo – over a sequence of four pilot studies conducted in both simulated and live classrooms – to draw teachers’ time and attention towards these students.

Conclusions and future work
In this paper, we have introduced Causal Alignment Analysis (CAA): a design framework for the data-informed design and iterative improvement of teacher awareness tools, linking the design of these tools to educational goals. We have illustrated the application and usefulness of CAA through a case study, demonstrating the iterative design alignment of a real-time teacher awareness tool with a hypothesized causal path from teacher tool use to student learning (Figure 4). The resulting prototype augments teachers’ awareness of student learning, metacognition, and behavior, while also measurably directing their time towards a subset of student processes that appear to have a negative influence on student learning outcomes.

While this case study may represent a step towards the design of teacher awareness tools that can measurably enhance student learning, it does not fully “close the loop” (Koedinger et al., 2013). To support iterative design, a CAA approach favors running larger numbers of small to mid-scale studies over running a single high-powered study. As such, it may not support strong causal inference. To better understand whether and how a teacher’s use of Lumilo influences student learning, we have recently conducted a larger-scale classroom experiment. Analyses of data from this experiment will enable us to investigate multiple hypothesized paths from teacher tool use to student learning (Figure 1), and thus to tease apart the distinct causal explanations that these paths represent. For example, although the analyses presented in this paper led to the improvement of Lumilo with respect to the hypothesized causal path pictured in Figure 4, it remains an open question whether the final link in this path (improved student learning) will hold in practice.

While the case study presented in this paper focused on data-informed design optimization with respect to teacher attention allocation across students (the top path in Figure 1), there are many other causal paths along which an awareness tool might be optimized. For instance, even if teachers are made more aware of critical moments, it may not always be clear how to effectively respond. Our design work with teachers suggests that they often desire more direct support (e.g., action recommendations) for planning and enacting effective interventions – especially in personalized learning contexts, where planning time can be very scarce (Holstein et al., 2018). A promising direction for future work may be to use CAA to explore whether and how an awareness tool could be designed to measurably enhance the effectiveness of teacher-student coaching interactions.

In summary, as the fast-growing research area of teacher awareness tools matures, we hope to see the design of these tools (within and beyond the academic Learning Sciences and Learning Analytics communities) increasingly guided by educational data and theory, in addition to user feedback. Causal Alignment Analysis provides a framework for making the goals and implicit assumptions behind the design of awareness tools explicit — in turn representing these assumptions as hypotheses to be continuously tested throughout a design process. Given the complexity of designing teacher awareness tools, we expect that such data-informed design approaches will be key to ensuring that they are not only useful and usable, but also beneficial for learning.

Endnotes
(1) In path analysis, the null hypothesis is that the estimated model is the true model. The p-value represents the probability, under the null, of observing a difference between the estimated and observed covariance matrices at least as large as the realized difference; a p-value above a given threshold (conventionally alpha = .05) implies a model cannot be rejected.
References

Acknowledgments
This work was supported by NSF Award #1530726, and IES Grant R305B150008 to CMU. The opinions expressed do not represent the views of NSF, IES or the U.S. ED. Special thanks to Gena Hong, Octav Popescu, Jonathan Sewall, Mera Tegene, Cindy Tipper, and all participating students and teachers.
The Effect of the Prior Collaborative Experience on the Effectiveness and Efficiency of Collaborative Learning

Jimmy Zambrano R., Universidad de los Hemisferios, Instituto Rumiñahui, jimmyz@uhemisferios.edu.ec
Femke Kirschner, Utrecht University, f.c.kirschner@uu.nl
Paul A. Kirschner, Open University of the Netherlands, University of Oulu, paul.kirschner@ou.nl

Abstract: This study investigates, from a cognitive load perspective, the effect of prior collaborative experience on the effectiveness and efficiency of collaborative learning. Performance, mental effort, and efficiency were measured during collaborative learning and in individual post-tests after one and seven days (i.e., retention and delayed retention test respectively). The results with 90 high school participants found that students who were members of experienced groups outperformed, invested less mental effort, and were more cognitively efficient than students in non-experienced groups in both tests. These results have important instructional implications for designing collaborative environments and provide support for the advantages of forming teams with relevant collaborative experiences before starting collaborative learning.

Keywords: collaborative learning, cognitive load theory, prior collaborative experience.

Introduction

Collaborative learning is an extensively used instructional technique in educational settings. It is the process by which individuals interdependently interact in small groups to learn from solving academic problems (Gillies, 2016; Slavin, 2014). This instructional approach has been broadly studied from multiple disciplines and perspectives (Hmelo-Silver, Chinn, Chan, & O’Donnell, 2013; Hmelo-Silver & Chinn, 2015). Although there is research from different perspectives using different theories, the conditions under which collaborative learning is effective and efficient are still insufficiently understood (Kester & Paas, 2005; Kirschner, Paas, & Kirschner, 2009). Studying collaborative learning from a cognitive perspective (i.e., taking human cognitive architecture into account) will provide valuable insights and guidelines for designing effective and efficient collaborative learning environments. The current study shows that in this context, having prior collaborative learning experience is an important determining factor.

References


Encountering and Becoming Role Models: Combating Underrepresentation in STEM

Leighanna Hinojosa, University of Colorado Boulder, Lehi6796@Colorado.EDU

Abstract: Neuroscience Education Outreach (NEO) is a program sponsored by a large public university located in the mountain west region of the United States which employs undergraduate students pursuing degrees in Science, Technology, Engineering, and Mathematics (STEM) to perform outreach and be liaison to community and university partners. Undergraduate instructors perform roles of preparing and teaching age appropriate lessons in local classrooms, thus brokering relationships among the university and K-12 students. Three case studies show how, through participation in outreach, undergraduate instructors from groups historically underrepresented in STEM are positioned as neuroscientists and gain professional skills which contribute to the development of their personal and professional identities in their STEM field. By using the Zone of Proximal Identity Development (ZOPID) framework, I analyze how identity development occurs incrementally within an authentic outreach community setting.

Keywords: Community Engagement, Outreach, Informal Learning, Identity

Research problem
There is a considerable lack of similar role models for women in Science, Technology, Engineering, and Mathematics (STEM) as (white) men make up the majority of STEM faculty in universities in the United States (National Science Foundation, National Center for Science and Engineering Statistics, 2013); and women are also more likely to leave STEM majors as compared to their male peers (Strenta, et al., 1994). Along racial lines, representation of women of color in STEM is further limited. According to the NSF (2013), White women comprised 20% of the STEM workforce, Black and Hispanic women made up 2% respectively, while Asian women comprised 5% of the STEM workforce and 1% of women who self-selected “other category” comprised the US STEM workforce.

This gender and racial disparity in STEM may signal to women they do not belong or cannot succeed in these fields (Walton & Cohen, 2007; Hermann et al., 2016). And because women entering STEM fields tend to underestimate their self-efficacy (Correll, 2001; Meece et al., 1982; Deaux & Emswiller, 1974), non-stereotypical role models, or historically underrepresented role models in STEM, have been thought to improve female students’ sense of self-efficacy and success in fields of STEM. In fact, in a psychological study examining effects of stereotypes, Cheryan et al. (2011) found differences in women’s sense of self-efficacy in STEM were mediated by the perceived dissimilarity from stereotypical role models. In addition, female role models have been shown to increase retention of women in STEM fields (Hermann et al., 2016); therefore, work on the effects of seeing one’s self as a STEM role model for women and minorities needs to be further explored.

Studies have also shown women and minorities in STEM are more likely to succeed in institutional environments with social and academic support systems such as mentorship opportunities (MacPhee et al., 2013). In a study conducted by Griffith (2010) using data from the National Longitudinal Survey of Freshmen (NLSF) and the National Education Longitudinal Study of 1988 (NELS:88) fewer women and minority STEM students persisted in their majors as compared to male and non-minority students, however differences in persistence also correlated with differences in college preparation and educational experiences. Institutional characteristics and opportunities for women of color in STEM to frequently engage their peers about course work, to be involved in STEM-related student organizations, and have access to participate in undergraduate research programs can positively impact female and minority student retention in STEM (Espinosa, 2011).

Finally, outreach programs have been shown to play a crucial role in the engineering education of graduate and undergraduate student instructors by fostering communication skills, leadership skills, and self-confidence among women who participate in delivering outreach (Pickering et al., 2004). Community outreach and engagement thus provides a perfect place to explore how participation in STEM related programs in an institution support undergraduate STEM education and professional identity development. By having nondominant students discuss their experiences, educational researchers can better understand how outreach is taken up by those students and what effect participation in outreach has on students and their identification with STEM. The purpose of this paper will be to examine, through case study, how individuals underrepresented in STEM participated in outreach, leading to their own development as “non-stereotypical” role models in STEM.
Furthermore, understanding how “non-stereotypical” role models develop can better aid in retention and engagement of groups underrepresented in STEM.

Scholarly significance

In this study, I tracked 10 students’ academic trajectories and observed students interacting with their peers and with K-12 youth. Preliminary data analysis suggests students in this program are learning leadership skills, being provided with networking opportunities, building self-efficacy in their respective fields, and going on to become professionals in STEM after graduation (Hinojosa, Torres, Callejas, & Speer, 2016). This paper describes how serving as a role model in outreach contributes to professional identity formation in First Generation College Students (FGCS), a group underrepresented in STEM and higher education. Key components of the student experience appear to be: encountering and being mentored by role models (program leaders) and peers (more experienced students) from underrepresented groups in their field of study, and then assuming the position of being a non-stereotypical role models themselves for youth of similar backgrounds. Understanding how non-stereotypical role models develop can better aid in engagement and retention of groups historically underrepresented in STEM by demonstrating to younger generations that scientists are as diverse as the people who populate this planet; thus, aiding to dispel negative stereotypes which may cause students to perform poorly in school (Marx, Ko, & Friedman, 2009).

The contribution of this study to the Learning Sciences community is to extend the theoretical framework of identity formation as a result of being mentored and mentoring others. The authentic outreach community works as a model for learning and identity development using 2 layers of mentoring and role modeling. In the context of the outreach program, NEO, the undergraduate instructors spend the fall semester preparing lessons and activities for outreach into K-12 local classrooms and developing their understanding of content knowledge in neuroscience topics. This is done with peers, the program leader, and guest faculty speakers doing current research in neuroscience. In the following semester—in the Spring—undergraduate instructors go out to local communities positioned as experts in neuroscience and as representatives of the university. Because program recruitment is focused on diversity, the majority of participants are women from underrepresented groups in STEM. This mean they are directly positioned not only as role models in neuroscience, but as non-stereotypical role models in that field. In this paper, I examine how this positioning of being a non-stereotypical role model contributes to identity development as a role model and as a professional in STEM.

Theoretical framework

The distance measuring the individual’s actual identity to a future possible identity imagined by self and others is represented by the Zone of Proximal Identity Development (ZOPID; Polman, 2006, p.246). Using the ZOPID—similar to Vygotsky’s (1978) Zone of Proximal Development (ZPD)—identity development is viewed as tied to the participant’s past positioning and positioning by self and others during social activities. Positioning within the outreach program is structured within a mentorship model which has two main components of mentorship: (1) mentoring from both faculty/staff; and (2) mentoring from more knowledgeable peers to newcomers in the program. A third component also emerges when undergraduate instructors go out to the classrooms as they are placed in the position of role models to youth from similar backgrounds, thus modeling to youth that a scientist can appear as they do. Figure 1 (below) is a diagram of how undergraduate instructors move between the different stages of the outreach program and how their positioning and identification changes in moments of time. The red highlights the portions of the outreach program where participant’s identity positionings are incrementally scaffolded to become more knowledgeable peers to newcomers as well as experts in neuroscience to the local communities. This contributes to the development of a professional STEM identity and a non-stereotypical role model in STEM. By using the ZOPID, I seek to unpack how this development occurs over time for three NEO instructors who identify as Latina women and the first generation to attend college in their families.

Neuroscience Education Outreach (NEO) is an outreach program sponsored by a large public university located in the mountain west region of the United States. The goal of NEO is to empower and engage K-12 students to take charge of their mental and physical health through neuroscience education. Lessons are designed to teach children and adolescents about their brain through a series of fun activities around lessons which demonstrate how the brain works. Undergraduate students in STEM serve as the instructors who prepare the lessons and activities, go into the surrounding schools and communities (in pairs or more), and facilitate the age appropriate lessons and activities (see figure 1 for NEO program structure).
Figure 1. The outreach model is structured to accommodate both newcomers into the program as well as more seasoned peers I refer to as More Knowledgeable Peers (MKP). In the fall semester timepoint 1 (T1) outreach instructors, who are recruited in a manner which is focused on diversity, prepared age appropriate lessons and activities to go out into the local classrooms. They are scaffolded by the program leader, who invites university neuroscientists to discuss current research, as well as by the MKPs. In the Spring semester, or timepoint 2 (T2), outreach instructors positioned as the experts go into local classrooms and facilitate the lessons and activities. By doing so they become MKPs themselves and assist other undergraduates in planning future sessions in the following Fall semester, or timepoint 3 (T3). I seek to explain identity development as the undergraduate mentors move through this model in real time.

Methods and data sources
Participants at the time of the study included 10 undergraduate instructors, however for the scope of this paper I elaborate on 3 instructors (Sarah, Zena, and Abby), who identify as first-generation college students (FGCS). Sources of data include annual surveys and interviews as well as peer and classroom observations of interactions conducted throughout 2 academic years: Fall 2015- Spring 2017.

Analytic Approach: Trajectories of Identification. I use sociocultural theory (Wertsch, 1998) to conceptualize learning environments as places where participants connect their past, present, and future selves, and in this way, make sense of who they are which further leads to how they participate in those environments. Using the trajectories of identification framework (Polman & Miller, 2010) I theorize participation and identity as dynamic and co-constructed via participation in social interactions and activities. To further analyze trajectories of identification, I looked at the parameters of prolepsis, positioning, agency, and scene.

Prolepsis occurs on the individual, cognitive level, and refers to how participants (and those they interact with) make sense of and act out connections between the participant’s past (prior lived experiences and memories), present, and their imagined future (Cole, 1996). This imagined future can be either imagined for the participant by others and/or by the participant him-/herself. Prolepsis therefore can lead the participant toward, or deter away from different types of imagined individual trajectories or pathways. The parameter includes “proleptic references” to the participants’ past (or similar types of people); to possible futures (for self and others similar); and finally, analyzes how present-day activities and interactions connect to either past or future selves. These proleptic references include stories which refer to role models and mentors, as well as imagined stories of themselves and others. Prolepsis can therefore be seen in the narratives participants tell about themselves and/or others, as well as seen in the actions and interactions of the individual participants during social activities. Prolepsis can also lead to positioning as identity is dialogically negotiated, and the enactment of envisioned identities and trajectories (both by self and others) shapes directly how that individual participates in that community. In the analysis, I distinguish between how one positions self and others, as well as how they are positioned by others within the outreach community. The individual can then either resist or accept forms of positioning which can be seen in social interactions and actions. How the individual reacts to positioning constitute forms of resistance to, or acceptance of certain trajectories of identification.

Positioning also refers to identifications from society and community. Society and community position individuals, either explicitly or implicitly, within that society before that individual begins to act in society. Even before one is born, there is an imagined future by parents (prolepsis) and one is positioned by culture and society (e.g., "girls’ clothes, toys, colors, etc.) to be something imagined for that person (Turner & Reynolds, 2010). Therefore, prolepsis and positioning are also shaped by societal factors and influences which further shape
trajectories of identification. These influences at the societal level may refer to identifications such as stereotypes (which are often negative, promoting a deficit view based on race, gender, etc.), or may refer to lack of representation or role models, and are recognized by individuals as they participate in social activities and interactions (Peteet et al., 2015; Steele, 1997; Clance & Imes, 1978). Therefore, identity is constantly being negotiated and renegotiated during interactions with others; and with how the individual is positioned by self, others, and also society. Agency refers to the capacity of an individual to act independently and be an agent in their own decision making (and thus identification and participation). In demonstrating competence in neuroscience content and outreach responsibilities, undergraduate instructors begin to position themselves as agents in their own trajectories of identification.

Lastly, learning environments are both structured and constrained by community and institutional goals and ways of positioning, referred to as scene. The STEM environment is not a-cultural, but instead STEM practices are embedded in deep cultural contexts which may be taken for granted as “norms” by insiders; and this is further embedded in a university environment. The community and institutional environment (and embedded norms) are referred to as the scene or place. Analyzing scene included looking at how particularly types of people and practices are taken up or not by the university outreach program. In conclusion, prolepsis, positioning, agency, and scene interact to form trajectories of identification for the individual leading to identity development.

Because identity is constantly evolving in time and is situated in context, I used narrative analysis to examine expressions of identity (de Fina, 2015) in multiple interviews trajectory of identifications framework. I examined interviews and field notes of interactions with public and peers to look for positioning cues and proleptic references. Salient social categories began to emerge. One particular category, first generation college student (FGCS), encompassed other traditionally underrepresented groups in STEM and presented elements of social influences (i.e., lack of role models or college preparation) contributing to that identification. Using the ZOPID (Polman, 2010) and trajectories of identification framework (Polman & Miller, 2010), I analyzed trajectories of identification in the development of non-stereotypical role models for three FGCS instructors as they progressed in their role in outreach and in their college career.

Background

Women and other minoritized groups are lumped together when NSF and other institutions discuss underrepresentation in STEM disciplines and submit proposals to aid in bolstering that representation, yet the term “underrepresented” is never explicitly clarified other than its inclusivity of women (including white women) and minorities (people of color). Through analysis of interview data and examining the role the category of FGCS plays, one can see members of FGCS tend to also share other categories which are considered underrepresented in science and engineering. For instance, Sarah, Zena, and Amber in this study share the categories of FGCS and Latina and have similar tensions and experiences when adapting to the college environment. In addition, these women of color find encountering and becoming role models exciting and inspiring for their own persistence in STEM. In fact, atypical of the current demographics of the university and its STEM programs, the majority of the instructors in NEO are women and/or minorities in the STEM field.

This type of environment is referenced by Amber as providing guidance and mentorship in academia, as well as providing a community of women in science with similar goals of outreach (i.e., a social and academic support system). In the case of Sarah, outreach to those with “similar backgrounds” is especially appealing as she begins to take on the identity of role model. Lastly, while the majority of the instructors are women from nondominant communities, there is also a mix of men and women from dominant communities.

Results

Examining FGCS more closely from the point of view of FGCSs, one can see the main associative factor is lacking knowledge of the college experience, lacking role models, lacking economic resources, and being a minority (and/or woman) student in STEM, which resonated in the narratives of Amber, Zena, and Sarah (Hinojosa, 2018). A first-generation college student is a student whose parents have not obtained a degree from a higher learning institution. This implies a lack of family resources or role models when students go into college. In all excerpts Sarah, Amber, and the program leader make connections to social factors which further contribute to attrition from STEM: being female, being from a lower SES background, being Latina, being a first-generation college student, as well as lacking role models in STEM. Sociohistorical factors influence the present-day scene in higher education and these factors contribute further to the academic achievement gap and STEM underrepresentation (Spring, 2016; Solomon, 1985). While this does not disprove any profound finding previously about underrepresentation in STEM, it does provide evidence outreach and community engagement is an effective medium to create more non-stereotypical role models of students who are underrepresented in science.
The NEO program structures itself as a mentorship model, and has two main components of mentorship: 
(1) mentoring from both faculty/staff; and (2) mentoring from more knowledgeable peers to newcomers in the 
program. A third component also emerges when undergraduate instructors go out to the classrooms as they are 
placed in the position of role models to youth from similar backgrounds, thus modeling to youth that a scientist 
can look like they do. Brooke (all proper names are pseudonyms) is the director of the outreach program, and is 
frequently referred to by the instructors as being a role model, resource and mentor. While NEO’s goal is to 
disseminate neuroscience information to the public, the instructors are receiving a level of role modeling and 
guidance which ultimately contributes to their persistence in STEM. Preparing instructors to teach lessons and 
facilitate activities is akin to an apprenticeship model whereby a novice is introduced to the content knowledge, 
and scaffolded on how to manage classroom interactions and activities by more knowledgeable peers (experienced 
instructors) (Rogoff, 1995) to practice science communication and public speaking skills. Lastly, Brooke positions 
the instructors as expected to succeed in their college careers, and as young scientists and role models.

Case Study 1:
Sarah is a double major in biology and philosophy in her third year at the university (after much uncertainty and 
switching majors), and currently works in a neuroscience and psychology lab on campus studying the effects of 
physical activity on the brain. She regrets not starting out in neuroscience, and attributes her decision to lack of 
knowledge about neuroscience before starting college. She is interested in pursuing a career in medicine yet has 
expressed uncertainty in her future career plans. Below she discusses neuroplasticity to high school female 
students and how first-generation students may struggle more in college compared to the traditional student (due 
to “inequalities”, perhaps referring to low SES or lack of knowledge/college preparation). She remarks how this 
was frustrating at first, but the brain can learn and adapt.

[Excerpt 1]: …I kinda spoke about my experience as a first generation college student how academia is—especially at the higher levels—is really challenging…[a]nd how often times you just wanna kinda say like…“Oh my gosh, I’m so dumb”, or “I’m so behind” but neuroplasticity kinda gives you evidence to suggest that you can improve inequalities that you’re not happy about… like studying habits and stuff and… neuroplasticity…helps you become better at something that you wanna do (Sarah, Interview, August 2016).

The above excerpt is from an interview with Sarah about her teaching experience in a college prep group 
for high school girls from low socioeconomic background, primarily from the local Latino communities—these 
students are on the college track and would be the first in their families to attend. Prolepsis was apparent when 
Sarah expressed her feeling “connected” to the students instantly because she holds membership in a social 
category the students are likely to share in their near future. Sarah also stated she was from a “similar background” 
and later stated she felt this connection due to the shared category of being Latina and low SES, as well as being 
a FGCS. Sarah noted that she saw herself as a role model to students (who reminded her of herself in high school) 
that day and even exchanged emails with one student interested in neuroscience. This went beyond the 
requirements of outreach as Sarah took a more personal interest in the imagined trajectories of the students with 
whom she interacted.

Case Study 2:
Amber graduated from the university with a degree in Biology in December 2016. Amber positioned herself as a 
FGCS during her college career. After she graduated she took a job as an analytical chemist in order to gain lab 
experience and go into the health field. Her desired career and projected identity is a diabetes researcher. She saw 
Brooke and other female professors as role models and found a supportive environment of peers in her outreach 
community.

[Excerpt 2]: I have a huge mentorship program because there's a huge, through [Brooke], through the other students, there's a huge mentorship part…(Amber, Interview, May, 2016).

This was in stark contrast to the university scene for Amber, as she lacked a sense of belonging at first, which she 
contributed to her dissimilarity from the average university student. Amber remarked that she could not quit school, 
despite typical FGCS challenges, because she wanted to be a role model to others, especially to young girls 
interested in science. As noted in the next case, Zena shared similar sentiments.

Case Study 3:
Zena was a psychology major (after much uncertainty and switching majors) in her fifth year at university, and wanted to pursue a career in education. She attributed her motivation to teach as tied directly to her participation in NEO. During outreach, she was positioned as a neuroscientist by the NEO director and by classroom teachers and students. Zena also used her social categories of FGCS and Mexican to connect to communities with people sharing a similar background. In this way, Zena was positioned as a non-stereotypical role model in the Latino communities she interacted with.

In Spring 2017, Zena organized events for the Latino community to involve parents to discuss educational neuroscience and how that related to childcare. Many of the parents and community members positioned Zena as a role model for their children, and were very interested in hearing the story of a FGCS succeeding. Parents began to see their own kids’ possible futures as being similar to Zena’s. Zena felt she could give the Latina mothers advice based on her own hardships and adversities as a FCGS, and in this way, was positioned by her community and herself as a role model of a successful university student. “Telling [the community] what I do I feel… gives them a broader sense of what they can do” (Zena, Interview, April 2017).

Zena continued to express a sentiment Amber expressed in her 2016 interview:

So like I mentioned before, letting other people down...’Cause you know I want to be an inspiration to someone... to at least one person and so being able to say, “yeah, I came here and it was hard and difficult, but I did it and look at where I am now” kind of thing. I want to be able to say that and the kids that we go and teach are like: “wow, you go to university, you go there! How did you make it that way?” And so if I didn’t, if I decided no, maybe I’ll go here instead cause it just feels safer or closer to whatever then I would feel like I kind of gave up on something and I was not what I thought I was (Zena, Interview, April 2017).

In the above excerpt Zena’s identity of being a successful first-generation college student is linked to her success at the university she attends, which is known for its excellence in STEM. By doing service work as a student representative of a major university, which lacks in diversity and in her experience, has had professors discourage her from pursuing her degree, Zena is combating those institutional and societal identifications and positionings imposed upon her in the university setting. She resists this scene by being a role model to others who might face similar challenges in pursuit of a higher degree at a major STEM university. Zena thus gains inspiration from outreach, as she inspires youth she interacts with, to continue her degree in this particular university: “...the kids I teach have been really inspiring to me ‘cause...I wanna work hard and fight [for them] (Zena, Interview, April 2017).” Themes of connecting to youth from similar backgrounds, seeing one’s self as a role model to youth and simultaneously building confidence and inspiration to continue in school are shared among Zena, Amber, and Sarah.

This bias of having lower expectation for poor students (and students of color) speaks to a larger deficit discourse of teachers, administrators, and policy makers (Benner & Mistry, 2007; Diamond et al., 2004; Cooper et al., 1982). In the U.S., where K-12 teachers are predominantly white and from middle class backgrounds, this can be devastating as classrooms are becoming increasingly diverse (Steele, 2001;2008). Furthermore, this assumption resonates with teachers from nondominant backgrounds as in the case of Zena. Zena sees this as important to her projected future trajectory as a K-12 teacher:

I guess, the assumption that like lower income kids don’t know as much...those were definitely challenged. I was like “oh my gosh, why have I been thinking this?” They definitely know as much and they are definitely interested and want to learn this stuff and...that was helpful and it really challenged me and totally changed my mind and now I’m like “oh my gosh, I cannot be thinking this way if I’m going to be a teacher” (Zena Interview, April 2017).

Zena specifically points out her challenged assumptions of low income students not knowing as much as students from more affluent groups (i.e., lower expectations of the knowledge lower income students bring to the classroom). This was a shock to Zena as she herself is from a lower income background, wants to work with those youth, and is on track to become an educator. These biases and assumptions are deeply embedded and can remain unconscious; only when those biases are challenged by our immediate interactions with those we don’t normally interact with do the biases come to the surface. The outreach instructors in NEO, including Zena, spent time after each lesson writing their reflections after performing outreach for both program feedback as well as for their own personal learning and growth. The reflexivity afforded by this practice may have contributed to her asserting agency in social awareness to overcome personal bias.

Discussion of major findings
By examining identity formation through moments in time within an authentic outreach model using the theoretical framework of the ZOPID, the practices associated with STEM retention and student achievement in their STEM degrees can be seen as encountering and becoming non-stereotypical role models. In the cases of Sarah, Zena, and Amber, these Latina women share the category of FGCS and have similar tensions and experiences when adapting to the college environment, or the university scene. The motivation behind doing outreach for all three participants was to empower communities with less access to pertinent neuroscience information to effect change in their own lives, and to inspire others to do science by being roles models to youth like themselves. Evidence shows the outreach instructors both found role models in their field (in the program supervisor and more knowledgeable peers), as well as saw themselves as role models during interactions with K-12 students. Thus, prolepsis may occur in the K-12 youth’s mind as they see themselves in the outreach instructors, leading to the development of trajectories of identifications involving science in the future for the youth. Hence, the distance between their imagined identity and their actual identity (the ZOPID) diminished. During outreach participation, the three instructors were positioned by themselves and others as scientists and role models for youth from nondominant communities interested in exploring or pursuing a science career. Some instructors explored other challenges in participating in outreach in nondominant communities, and upon further examination, a larger negative stereotyping discourse of low income, non-white communities can be seen in the educational scene. Zena mentioned this tension even though her own background is low income, non-white, and first generation to go to college. As noted elsewhere, unconscious biases related to students' membership in minoritized groups can be harmful for students as teacher expectations affect student outcomes (Benner & Mistry, 2007; Diamond et. al, 2004; Cooper et al., 1982). Using reflexivity, Zena began to draw connections from her assumptions and actual interactions with youth from nondominant communities, thus allowing her more agency in her conceptions of low income youth. This allows her to see youth in new imagined trajectories, leading to a different positioning of the youth which links their pasts to present and ultimately to imagined futures of being a successful college student in STEM. This provides room for a rippling effect of role models begetting future role models.

Conclusion and further implications

This study provides evidence of how mentorship models within university outreach can be an effective site for the development of non-stereotypical role models in science. In order for Amber, Zena, and Sarah to continue with their college careers they all had to overcome the scene of the university environment and how those in the environment positioned them differently than how they were positioned during outreach. For example, professors intimidated the FGCSs during classes often remarking “you should already know this,” referring to domain content knowledge, leaving Zena and others to equate their lack of college preparedness to lack of intelligence. During outreach, instructors found a group of similar and supportive peers, as well as found female role models in science who positioned them also as such. The three FGCS instructors have since learned how to talk to faculty members (networking) and how to navigate STEM institutions contributing to their engagement and retention in STEM. Zena and Amber attribute their networking skills to having a strong social network of peers, mentors and non-stereotypical role models. This impact, i.e., empowering youth to effect change in their lives, is something all instructors pointed to as their personal motivation to perform outreach work. This study thus demonstrates personal growth and professional identity development during participation in community outreach. Future implications involve bringing this model and framework to university settings to foster student success and bolster diversity in STEM higher education and ultimately into STEM professional fields.

References


Acknowledgements
This project was supported by the Office of Outreach and Engagement and School of Education at the University of Colorado Boulder. The content is solely the responsibility of the authors and does not necessarily represent the official views of the university.
The Influence of Students’ Cognitive and Motivational Characteristics on Differences in Use and Learning Gain in an E-Learning Environment

Charlotte Larmuseau, Jan Elen, Piet Desmet, Fien Depaepe;
charlotte.larmuseau@kuleuven.be; jan.elen@kuleuven.be; piet.desmet@kuleuven.be; fien.depaepe@kuleuven.be
KU Leuven

Abstract: Differences in use of e-learning components can be influenced by student-related characteristics. In this study, an e-learning environment was developed in line with the four component instructional design (4C/ID) model. The four components consist of authentic problem-based tasks (i.e., learning tasks), various support (i.e., supportive and procedural information) and additional drill-and-practice exercises (i.e., part-task practice). This study firstly investigates the influence of students’ prior knowledge, task value and self-efficacy on students’ use of the components. Secondly, it examines the influence of students’ use of the components on their learning gain, taking into account their characteristics. Results of 161 students reveal a significantly negative influence of prior knowledge on students’ use of learning tasks and part-task-practice, whereas task value has a significantly positive influence on use of learning tasks and supportive information. Furthermore, use of learning tasks, procedural information and students’ prior knowledge significantly contribute to students’ learning gain.

Keywords: instructional design, e-learning, distance education, prior knowledge, motivation

Introduction

Advances in e-learning, such as virtual and asynchronous self-paced approaches, have increased learners’ autonomy. Consequently, e-learners can control when, what and how to study, and accordingly students have to self-direct their learning (Joo, Lim & Kim, 2013). An instructional design model that amplifies self-directed learning is the 4C/ID-model (van Merriënboer & Sluijsmans, 2009). A 4C/ID-based learning environment contains four different components: (1) concrete, authentic, problem-based, whole-task experiences (i.e., learning tasks), (2) additional drill-and-practice exercises (i.e., part-task practice) and two categories of support such as (3) background theory (i.e., supportive information) and (4) just-in-time information (i.e., procedural information). In this study, an e-learning environment was developed in line with the 4C/ID-model. The additional exercises and various support were non-embedded. More specifically, this implies that on the learner’s initiative, additional part-task practice and/or support could be selected. Learning tasks were less optional, since students were strongly encouraged to consult them. Nevertheless, solving all of them or making various attempts to solve a learning task was optional. Consequently, we expected that one student would quickly proceed from learning task to learning task, while another student would select part-task practice or consult supportive information. Accordingly, in this context, self-directed learning refers to students taking initiative in diagnosing their learning needs by identifying appropriate resources (i.e., one of the four components), evaluating their outcomes (i.e., scores on learning tasks) and implementing learning strategies (e.g., consulting additional exercises or support in order to improve their results). Furthermore, prior research indicates that using different e-learning components can be influenced by cognitive and motivational characteristics (Greene & Azevedo, 2007; Jiang, Elen & Clarebout, 2009; Rienties, Tempelaar, Van den Bossche, Gijselaers & Segers, 2009). Therefore, the first aim of this study is to investigate the influence of students’ cognitive and motivational characteristics on students’ use of the four components of a 4C/ID-based e-learning environment. In accordance with prior research, students’ cognitive (i.e., prior knowledge) and motivational (i.e., task value and self-efficacy) characteristics are included in the research model. Moreover, the provision of different components that provide the opportunity to consult additional exercises and various support can enhance students’ performance (Lust, Juarez Collazo, Elen & Clarebout, 2012). By specifically looking at the use of each of the four components, more insight can be gained in how students’ use of a specific component contributes to students’ learning gain. Therefore, a second aim of this study is to measure how students’ use of the four components of the 4C/ID model, influence students’ learning gain, taking into account their cognitive (i.e., prior knowledge) and motivational (i.e., task value and self-efficacy) characteristics. Accordingly, to achieve both aims, a structural model is suggested that integrates students’
cognitive and motivational characteristics, the four components of the 4C/ID-model and students’ learning gain in order to elucidate the relationships among these variables.

**Theoretical background**

An instructional design model that stresses integration and transfer of learning is the 4C/ID model elaborated by van Merriënboer (1997). The 4C/ID model is acknowledged as one of the most effective models for designing learning environments that facilitate the acquisition of integrated sets of knowledge, attitudes and skills (Merrill, 2002). The basic concept of the 4C/ID model is that learning environments can be described in terms of four interrelated components: (1) learning tasks, (2) part-task practice, (3) supportive and (4) just-in-time information. The learning tasks are concrete, authentic, problem-based, whole-task experiences. Learning tasks are grouped in task classes and sequenced based on their degree of difficulty in order to prevent cognitive overload for the learners, as this could hamper learning and performance (van Merriënboer & Sluijsmans, 2009). Support is provided in two distinct manners, that is, supportive and procedural information. Supportive information is basically, the theory and therefore supports the learning and performance of the non-recurrent, problem solving and reasoning aspects of learning tasks. It helps learners to link the presented information to existing schemata, that is, to what they already know in order to solve the learning tasks. Procedural information is prerequisite to the learning and performance of recurrent aspects of the learning tasks in each task class. It allows students to complete and learn routine aspects of learning tasks by specifying exactly how to solve the routine aspects of the tasks. It is presented just in time when learners need it. Furthermore, part-task practice supports the more complex whole task learning by providing additional exercises for selected recurrent constituent skills (van Merriënboer, 1997).

A 4C/ID-based e-learning environment is claimed to stimulate self-directed learning by providing different components at the student’s disposal (van Merriënboer & Sluijsmans, 2009). Moreover, by giving students control of the use of the different components, differences in use based on the students’ learning needs can be possible. Former research indicated that differences in use of e-learning components is influenced by learner-related characteristics (Greene & Azevedo, 2007; Jiang et al., 2009; Rienties et al., 2009). An important learner characteristic that influences differences in use is students’ prior knowledge. Cognitive load theory suggests that students with low prior knowledge cannot immediately be confronted with highly difficult learning tasks. Accordingly, students’ cognitive load can be reduced by consulting support and guidance (van Merriënboer & Sluijsmans, 2009). This implies that students who find the task difficult, benefit from various support and additional exercises in order to improve their performance. Nevertheless, selecting various support or making additional exercises can be very challenging (Clarebout, Horz, Schnotz & Elen, 2010; Lust et al., 2012). This requires awareness of learning gains, and therefore includes metacognitive monitoring of knowledge and comprehension, which might be difficult to achieve when students experience high cognitive load. As a higher level of prior knowledge reduces cognitive load during e-learning, students with higher prior knowledge might face lower cognitive load and accordingly, are more likely to cope with these metacognitive requirements compared to students with lower prior knowledge. Subsequently, based on cognitive load theory, differences in use can be influenced by the level of students’ prior knowledge (Moos & Azevedo, 2008; van Merriënboer & Sluijsmans, 2009). Despite this theoretical claim, several studies that investigated the impact of students’ prior knowledge on differences in use did not find any effect. Van Seters, Ossevoort, Trampler and Goedhart (2011) used e-learning materials to demonstrate how 94 university students work differently based on their prior knowledge. They measured the learning paths students followed when working with adaptive e-learning material. The learning path was determined by average step size chosen, average number of tries, total number of exercises and time needed to finish. They found that differences in students’ prior knowledge did not have an effect on students’ learning paths. Taub, Azevedo, Bouchet and Khosraavifar (2014) studied differences in use of an e-learning environment in relation to their prior knowledge. Results of 112 undergraduates revealed that all students visited similar number of relevant pages regardless of their level of prior knowledge. Jiang et al. (2009) conducted a study where they measured variety in non-embedded (i.e., optional mode) tool use in an e-learning environment (e.g., checklist tool, information list, calculator etc.). Tool use was measured by frequency of tool use and proportional time spent on tools. Results of 58 bachelor students revealed that there was no influence of prior knowledge on the quantitative aspects of tool use. Aforementioned studies seem to confirm that students do not grasp learning opportunities based on their level of prior knowledge (Lust et al., 2012).

In contrast to prior knowledge, empirical evidence revealed that motivational characteristics have an important influence on students’ learning behavior in e-learning settings (Chen & Jang, 2010; Rienties et al., 2009). According to expectancy-value theory, self-efficacy and task value are two key components for understanding students’ specific use and academic outcomes. Self-efficacy is defined as a learners’ ability to execute the required behavior necessary for success (Greene & Azevedo, 2007). There is evidence that self-
efficacious students participate more readily, work harder and persist longer when they encounter difficulties than those who are uncertain about their capacities (Zimmerman, 2000). Task value essentially refers to the reason for doing a task. More specifically, students with high task value pursue enjoyment of learning and understanding of new things (Joo et al., 2013). Martens, Gulikers and Bastiaens (2004) investigated the impact of task value on the use of an e-learning environments. The participants were 33 higher education students. Results showed that students with high task value did not do more, but did other things than students with low task value. Analysis of log files showed that students with high levels of task value showed proportionally more explorative study behavior. The explorative pages were defined as pages that students were not explicitly directed to by the external source. Studies also indicate relationships between self-efficacy, task value and performance. Bong (2001) conducted a path analysis to investigate the relationships between task-value, self-efficacy and enrollment intentions i.e., use of the e-learning environment in an online learning context. Results of 168 undergraduate university students showed that task value was linked to course enrollment intentions. However, no influence of self-efficacy on course enrollment intentions was found.

Furthermore, differences in use of e-learning components can also influence students’ learning gain. Lust et al. (2012) conducted a literature study which provided empirical evidence for the beneficial influence of differences in tool use i.e., information, processing and scaffold tools, on students’ learning gain. In addition, former research indicated that student-related characteristics can directly influence students’ learning gain. Song, Kalett and Plass (2016) studied the direct and indirect effects of university students’ prior knowledge, task value and self-efficacy on students’ learning gain in an e-learning environment. SEM revealed that university students’ prior knowledge directly positively affected their learning gain, but no significant effects of task value and self-efficacy on students’ learning gain were found. By contrast, Joo et al. (2013) investigated 897 learners in an online university. Using SEM they found significant positive relationships between both task value and self-efficacy on the final grade on the course. These studies indicate, that when we measure the influence of differences in use on students’ learning gain, we should control for student-related characteristics. Summarized, based on these aforementioned theoretical and empirical claims we hypothesize that (1) prior knowledge, self-efficacy and task value can have an influence on students’ use of the four components and (2) that students’ use of the four components can be beneficial for students’ learning gain, taking into account student-related characteristics (i.e., prior knowledge, self-efficacy and task value). Consequently, we formulate following research questions:

- **RQ1:** How do students’ cognitive (i.e., prior knowledge) and motivational (i.e., self-efficacy and task value) characteristics influence use of the four components of the 4C/ID-based e-learning environment?
- **RQ2:** How does differences in use of the four components of a 4C/ID-based e-learning environment influence students’ performance, controlled for students’ cognitive and motivational characteristics?

**Methodology**

**Measurement instruments**

The e-learning environment in the present study focuses on French as a foreign language. It contains four learning tasks and takes about 1 hour and 15 minutes to complete. The level of difficulty was aligned with the level that students in the Flemish part of Belgium are expected to reach at the end of the secondary school, i.e., level B1 of the Common European Framework of Reference. Participants were 161 first year Psychology and Educational Science students. The majority of the students were female (91%). The average participant was 20 years old ($SD = 2.92$). Participation to research is part of the students training program, but French was not a part of their training program.

The administration procedure of the study consisted of two administration sessions. The first administration session started with a pretest, an introduction of the e-learning environment and an additional questionnaire on task value and self-efficacy. Task value and self-efficacy were both measured after the introduction of the e-learning environment to make sure students had sufficient insight into the e-learning content to form an opinion. The students were asked to use the e-learning environment at home during two weeks. As the learning content was not a part of their educational program, they received the instructions that consulting the four components was optional and that there was no strict trajectory to follow, but that consulting the learning tasks was strongly recommended. After the intervention of two weeks a second administration session took place where students received a posttest.

The learning environment is developed along the instructional design principles of the 4C/ID model. The first component deals with the learning tasks in the e-learning environment. These learning tasks are based on authentic situations for instance, ordering food in a restaurant. The learning tasks were sequenced in an easy-to-
difficult order, and were clustered in a task class. At the end of the task class, students must be able to have a fluent conversation at the restaurant. Students receive automatic generated feedback based on their scores. In order to have a fluent conversation, students must master the grammar (e.g., l’article partitif), vocabulary (i.e., food), skills (i.e., listening) and attitudes (i.e., elementary courtesy in a restaurant). Accordingly, during these learning tasks students can click on links to consult one of the other three components of the 4C/ID model. More particularly, students can consult additional exercises i.e., part-task practice (e.g., drill-and-practice exercises to practice food vocabulary) and support, respectively, supportive (e.g., grammar explained by theory) and procedural information (e.g., grammar explained by using keywords). Subsequently, the supportive, procedural information and part-task practice are non-embedded i.e., they are at the disposal of the students but the students can decide whether or not to use them. Learning tasks were partly non-embedded, since students were free to make as many learning tasks (e.g., several attempts) as they wanted. Nevertheless, as aforementioned, they were also partly embedded (i.e., less optional) since students were strongly advised to complete the learning tasks during the first administration session and in addition learning tasks were clustered in a task class (i.e., predefined order).

Measurement instruments
To measure the learning content a quantitative paper-and-pencil instrument on French was used as pretest (i.e., prior knowledge) and posttest (i.e., students’ learning gain). The instrument consists of 60 items and focuses on knowledge (i.e., grammar and vocabulary) and skills (i.e., listening, writing a conversation). The level of difficulty of the test was B1 of the Common European Framework of Reference (Evens, Elen & Depaepe, 2017). The instruments’ reliability was explored by calculating internal consistency i.e., Cronbach’s $\alpha = .90$ for the pretest and Cronbach’s $\alpha = .89$ for the posttest (Cueiford, 1965).

Within this study the constructs self-efficacy and task value were derived from the Motivated Strategies for Learning Questionnaire (MSLQ; Duncan & Mckeachie, 2005). For this study we used the constructs self-efficacy (5 items) e.g., “I expect to do well in this course”, and task value (6 items) e.g. “It is important for me to learn the course content”, of the motivation section. The questionnaire had a 7-point Likert-type response format with values ranging from strongly agree (7) to strongly disagree (1). The questions were translated into Dutch. Construct validity was checked by conducting a confirmatory factor analysis (CFA). The factor loadings from the latent variable constructs were all significant, had standardized values ranging from .74 to .93 and an average variance explained (AVE) of .76 for self-efficacy and .62 for task value. This verifies that the two measurement models of self-efficacy and task value were each measured well in the current data (Khine, 2013).

Internal consistency was investigated by measuring the Cronbach’s Alpha. The Cronbach’s Alpha for self-efficacy was $\alpha = .94$ and for task value $\alpha = .84$, which indicates high reliability (Cueiford, 1965).

Information of students’ use of the four components of the 4C/ID model was collected by tracking students’ activity, i.e., registration of views and interaction by the Moodle learning management system (LMS) during two weeks for each component. All data were collected on an aggregate module level and afterwards merged, based on the use of the different components. All data were anonymized through means of the use of random codes to safeguard the identities of the students.

Results
Students had an average of 50.97% ($SD = 18.17$) on the pretest and an average of 64.59% on the posttest ($SD = 15.88$). The average student replied “neutral” in terms of motivation ($SD = 1.13$) and self-efficacy ($SD = 1.10$). An overview of the distribution of the use of the four components (i.e., activity tracked by Moodle LMS and the amount of students that used the different components) can be found in Figure 1. The average time spent on using the online learning is 66 minutes ($SD = 27.34$, min. $= 10.44$ minutes, max. $= 151.43$ minutes or approximately 2.5 hours).
Structural equation modeling (SEM; \( N = 161 \)) was conducted in order to firstly, investigate the influence of students' cognitive (i.e., prior knowledge) and motivational (i.e., task value and self-efficacy) characteristics on the use of the components. And, secondly, to investigate the influence of the use of the components on students’ learning gain, controlled for student-related characteristics (i.e., prior knowledge, task value and self-efficacy). For the missing values a two-stage approach was applied. This approach obtains a saturated maximum likelihood (ML) estimate of the population covariance matrix and then uses this estimate in the complete data ML fitting function to obtain parameter estimates (Savelei & Bentler, 2009). Lavaan (Rosseel, 2012) converged normally after 59 iterations. The hypothesized model, provided an adequate fit to the given data \( \chi^2/df = 155.56/99 = 1.4; \ SRMR = .05, \ RMSEA = .06, \ CFI = .96, \ TLI = .95 \). The \( \chi^2 \)-test indicates the difference between observed and expected covariance matrixes and should be non-significant. However, \( \chi^2 \)-test is highly dependent on sample size and therefore normed \( \chi^2 \)-test is often considered i.e., \( \chi^2 \)-test divided by the degrees of freedom. Values smaller than 2.0 are considered to indicate acceptable fit (Kline, 2013). In addition to \( \chi^2 \) statistics, the root mean squared residual (SRMR), the root mean squared error of approximation (RMSEA), comparative fit index (CFI) and the Tucker-Lewis Index (TLI) were examined. SRMR is the difference between the observed variance and the predicted variance. A value less than .06 is considered a good fit. RMSEA adjusts for the complexity of the model and the size of the sample. The value for acceptance is >.05. A value of CFI and TLI between > .95 indicates good fit. Assessing all measures and considering the above statements, the original structural model was accepted and considered adequate (Khine, 2013).

Figure 2 gives an overview of the results of the research model. Significant influences of students’ characteristics were found on the use of the components, more specifically, students’ task value influences the use of learning tasks (\( \beta = .21, p < .05 \)) and supportive information (\( \beta = .22, p < .05 \)). No significant relationships were observed between students’ self-efficacy and the use of the four components of the 4C/ID-model. A significant negative relationship was found between prior knowledge and part-task practice (\( \beta = -.21, p < .05 \)). The variance explained for the dependent variables was \( R^2 = .07 \) for learning task, \( R^2 = .03 \) for part-task practice, \( R^2 = .08 \) for supportive information and \( R^2 = .02 \) for procedural information. RQ2 investigated the influence of students’ use of the four components of the 4C/ID model on students’ learning gain, controlled for students’ prior knowledge, task value and self-efficacy. Results reveal a significant influence of the use of components of the 4C/ID-model on students’ learning gain, more specifically, the use of procedural information (\( \beta = .08, p < .05 \)) and learning tasks (\( \beta = .12, p < .01 \)) have a significant influence on students’ learning gain. Prior knowledge had a major influence on students’ learning gain (\( \beta = .91, p < .001 \)). The variance explained for students’ learning gain was \( R^2 = .79 \). In conclusion, students’ differences in use of the components is influenced by students’ task value and prior knowledge. Students’ learning gain is influenced by differences in use of learning task and procedural information. Additionally, students’ learning gain is mainly influenced by students’ prior knowledge.
Discussion

The current study strived to find evidence (RQ1) for the influence of students’ cognitive and motivational characteristics on students’ use of the four components and (RQ2) for the influence of differences in use of the four components on students’ learning gain, controlled for students’ characteristics. All variables were incorporated in one structural research model. Our results are based on data from a pre- and posttest (i.e., prior knowledge, students’ learning gain), questionnaires (i.e., task value and self-efficacy) and platform log data from 161 students. Results indicate that prior knowledge and task value induce differences in use of the components of the e-learning environment. More specially, prior knowledge has a significant negative influence on the use of learning tasks (e.g., authentic problem-based exercises) and part-task practices (e.g., drill-and practice exercises). Part-task practices contain additional exercises with more recurrent content in order to prepare students to solve the learning tasks which contain more non-routine content. Accordingly, results indicate that some students seem to be aware that they are lacking routine knowledge to solve the learning tasks. Furthermore, results indicate that students attempted the learning tasks a few times in order to obtain better scores (i.e., differences in activity for the learning tasks). These findings are in contrast to the study of Taub et al. (2014) indicating that there were no significant differences in students’ use of the e-learning environment (i.e., defined by the number of relevant pages visited) between lower and higher prior knowledge groups. Results are also different from the study of Jiang et al. (2009) which revealed no association between prior knowledge and the frequency of tool use and/or proportional time spent on tools. A possible explanation for the difference between our results and former studies might be that the uncontrolled setting in our study allows for more self-regulation. The study of Taub et al. (2014) and Jiang et al. (2009) were conducted in a controlled setting i.e., students worked with the e-learning environment under supervision. By controlling the setting students might feel the pressure to complete the tasks in a given time. As consequence students might follow a more traditional linear path. These differences in learner control can have an influence on students’ use of different tools. Additionally, results indicate that differences in use are influenced by students’ task value. Students’ task value seems to have a positive influence on the use of the learning tasks. This could imply that students put more effort in solving the learning tasks qualitatively. Students’ task value also has a significant influence on the use of supportive information (i.e., background information and theory in order to understand the content of the learning tasks in its entirety). This corresponds with the study of Martens et al. (2004) who also analyzed log files and found that students with high intrinsic motivation (e.g., task value) showed more explorative study behavior. A difference in the study of Martens et al. (2004) was that the variable exploration was calculated by dividing the number of explorative pages a student had visited by the total number of visited pages. Finally, self-efficacy had no influence on differences in use. These findings are consistent with the study of Bong (2002). A feature of self-efficacious students is that they work harder and persist longer when they encounter difficulties. As there was no influence, this could indicate that the tasks in the e-learning environment were not complex enough for the students to encounter difficulties (Zimmerman, 2000). Findings could also be due to the way in which usage was measured (i.e. course activity and enrollment intentions), as persistence is a feature that it might be possible that the spent time could give more information about this influence.
In this study, RQ2 investigated the influence of students’ use of the components on students’ learning gain, controlled for cognitive and motivational characteristics. Results indicate that use of the learning tasks and procedural information significantly contributed to students’ learning gain. These results imply that the combined use of learning tasks and procedural information influences students’ learning gain. The procedural information in this context is just-in-time information. Therefore, consulting this particular support and guidance can prevent learners from paying attention to irrelevant task aspects and subsequently reduce cognitive load, which on its turn can improve their task performance (van Merriënboer & Sluijsmans, 2009). Regardless these results, students’ prior knowledge still has a primary significant influence on students’ learning gain. This can be related to cognitive load theory, as the online course used rich learning tasks, which are based on real-life situations, these tasks can be highly complex. A risk of this approach is that the cognitive load imposed by the learning tasks is often excessive for students with low prior knowledge and may seriously hamper learning (van Merriënboer & Sluijsmans, 2009). Furthermore, this major influence of students’ prior knowledge could also be influenced by the study design, more specifically, the fact that a rather short intervention, spread out over two weeks, is probably not enough to exert a major influence on students’ learning gain.

Irrespective of the added value of these findings, some limitations in the current study should be mentioned. Firstly, this study gives information of what is used, but little information is provided on why students used these specific components. For instance, combining objective information with think-aloud protocols can give more insight in the actual cognitive processes (Winne, 2010). Secondly, by looking at the isolated use of the four components, little is known about how the different components are used. By analyzing log-data more in detail (i.e., looking at the sequences of use of the four components), more insight could be given on effective use. A possible example of effective use would be that when a student has an insufficient score, the student decides to consult supportive information. A third important limitation concerns the course design of the e-learning environment. Findings of the current study indicated that the use of learning tasks did not differ based on students’ level of prior knowledge. As the learning tasks were sequenced in a simple- to -complex order, were clustered in a task class and had a predefined order, this course design probably influenced these results. It might have been more interesting if students had been able to select learning tasks themselves (e.g., based on their level of difficulty). Follow-up studies should enable students to select their future learning tasks at their option. This would provide more information about the self-directed learning (i.e., evaluating their learning outcomes and selecting learning tasks based on their performance; van Merriënboer & Sluijsmans, 2009).

**Conclusion**

This study firstly provides more information about the influence of students’ motivational and cognitive characteristics on differences in use of the four components in a 4C/ID-based learning environment. Results indicate that students’ characteristics do influence differences in use of the four components when students receive a lot of learner control. Moreover, students seem to slightly adapt their behavior based on their specific learning needs and interests. Accordingly, more comprehensive information (e.g., theory) might challenge more motivated students. Additionally, students with lower prior knowledge seem to consult part-task practice to help them to reach a very high level of automaticity for selected recurrent aspects of real-life problem-solving tasks. As a result, findings indicate that the 4C/ID-model is an instructional design model that allows for a lot of learner control and therefore supports self-directed learning, by providing four components containing different information (or a different format of information) that can be consulted freely in an non-linear trajectory. Furthermore, results reveal the importance the combined use of learning tasks and procedural information on students’ learning gain, when controlling for students’ characteristics. These students’ characteristics amongst which prior knowledge, seems to have had a strong impact on learning gain. Therefore, more research should verify the impact of students’ characteristics on differences in use and learning gain, using more extensive interventions. Moreover, students’ cognitive load should be included to measure the impact of the complexity of the learning tasks. Overall, insight into differences in use based on student-related characteristics is an important step from an instructional design perspective. This could provide important suggestions for intervening and adapting the online learning environment to students’ learner-related characteristics and associated learning behavior.

**References**


Authentic Problem-Based Learning With Augmented Reality

Todd Ogle, Virginia Tech, todd.ogle@vt.edu
David Hicks, Virginia Tech, hicks@vt.edu
Aaron Johnson, University of Nebraska-Lincoln, ajohnson147@unl.edu
Thomas Tucker, Virginia Tech, thomasjt@vt.edu

Abstract: This work resulted in an authentic problem based – and place based – historical inquiry unit utilizing augmented reality and an engagement-first approach to scaffold disciplined inquiry, facilitate student understanding of change over time, and motivate and sustain students through the inquiry arc. The design was guided by cognitive research that recognizes both the psychological and social aspects of learning. The outcomes represent both research findings across 14 participating classes as well as a more specific look at the activities of one student group, documenting their experience and perspectives in-situ. Data gathered include interviews, usage logs, student notes and student final projects. All interviews and video recordings were transcribed, indexed and underwent inductive content analysis to identify key themes around motivation to learn and explicit learnings and negotiations with the app. Our findings suggest that AR can facilitate and sustain an inquiry through making the invisible visible in authentic and challenging settings.

Keywords: augmented reality, place-based learning, historical inquiry, authentic problem-based learning

This paper reports on the outcomes of an augmented reality (AR) application designed to scaffold schoolchildren’s investigation of the hidden local history of a former African American school – Christiansburg Institute. A transdisciplinary team of Virginia Tech faculty in History, Education, Computer Science, and Visual Arts, public school teachers, the school division’s social studies supervisor, and Christiansburg Institute alumni worked together to build the place based/problem based historical inquiry unit of instruction that used AR to allow students - with little or no knowledge of what the historical site was – to engage in cultural fieldwork. The app, CI Spy, was part of a unit designed to introduce 5th grade students to the site’s history, purpose, and significance while engaging with a range of historical sources in a place-based inquiry (See Johnson et. al 2017). At the initiation of this project we hypothesized that AR with a built-in explicit strategy/scaffold for historical source analysis (SCIM-C: Summarize, Contextualize, Infer, Monitor and Corroborate) could support in-situ analysis of various historical sources while leveraging the power of context and place to motivate and sustain students within and through a historical inquiry as they learned about school segregation. Currently, the site of Christiansburg Institute, once a thriving 13 building campus of over 180 acres, is little more than a single derelict building in an empty field next to an industrial park. Few in the community know of the school, and it is truly a hidden, difficult history that the school division is now seeking to integrate within the history curriculum.

We created CI Spy to make the invisible past visible to students. That is, CI Spy was designed to help students see the growth and demise of the campus by using virtual buildings to provide a sense of its physical scale and presence. Additionally, CI Spy allowed students to “virtually” enter into the lone inaccessible derelict classroom building (Figure 1) and two of the virtual representations of buildings - long since demolished – (Figure 2) in order to collect and analyze historical sources (Figure 3) as part of a student-centered inquiry into the local history of school segregation.

Figure 1. An inaccessible building (a) can be explored (b) with a virtual classroom via AR (c).
Figure 2. Students tap a virtual building (a), confirm they’d like to enter (b), and explore the evidence within (c).

Without the use of AR this fieldtrip experience, designed to foster an inquiry and also facilitate a sense of walking in the footsteps of others, would have been impossible. A sense of place, scale, and what once existed on the now abandoned site were provided via the app, while at the same time it integrated the explicit scaffold based on SCIM-C to support the students’ inquiry.

Figure 3. Students use the scaffold (a), explore a virtual room (b), and work in pairs to record field notes (c).

Historical inquiry, explicit scaffolds and the importance of place

The teaching of history has been observed by generations of students and researchers as clinging to a very specific pattern of teaching: the teacher talks and students listen, read, and answer questions in textbooks. Students are then expected to memorize facts and details that for the most part are “removed from their intrinsically human character” (Goodlad, 1984, p. 212) and completely at odds with visions of powerful teaching and learning elucidated by the field. What is often lost or forgotten in such teaching is a clear goal of explicitly preparing students to actively engage in inquiry, perspective taking, and meaning making. Historical inquiry is the process of engaging in purposeful and reflective mental activities that strategically explore multiple perspectives through the reasoned drawing of inferences, the integration and synthesis of information, the evaluation of reliability and perspective, and the generation of possible understandings and interpretations (Bain, 2005; Levstik & Barton, 2005; Wineburg, 2001). History learning as a specific domain requires deep inquiry; considering and evaluating historical sources brings disciplinary rigor into the process and helps children develop the best possible arguments for whatever stories are formed. Further, the process of considering different pieces of evidence from varying sources can help elicit an understanding that alternative stories exist with varying perspectives, and the stories with the most support may not necessarily be the stories we would prefer to tell. The recognition for perspectives affected by critical inquiry can help students to achieve a deeper understanding of the importance of according people in the past the same respect as we would want for ourselves (Lee, 2011).

Ongoing research on explicit scaffolds in history education and beyond has established that they can be effective in helping learners to understand the inquiry process and to analyze historical sources (Belland, Kim, & Hannafin, 2013; Hicks & Doolittle, 2008; Hicks et. al 2016). Currently there is only a small body of empirical research designed to examine how digital technologies can support the teaching and learning of the doing of history (Britt et al., 2000; Saye & Brush, 2002). Much of this research (Saye & Brush, 2002; Hicks, Doolittle & Ewing, 2004; Hicks & Doolittle, 2008) has sought to examine how multimedia tools can broadly scaffold the analysis of multiple and single historical textual sources as part of the process of engaging in problem based/historical inquiry.

Visitors to historic locations can experience intense engagement, a loss of the feeling of the passage of time, and a connection with others across time and space (Cameron & Gatewood, 2000; Maines and Glynn, 1993). This suggests that placing learners within real historic contexts where they can interact with people, material
objects, visual and spatial information (Schwartz & Heiser, 2006), and processes and scaffolds could enhance the quality of learning (Barron & Dobbs, 2015; Coulter & Polman, 2004; Dunleavy et al., 2008; Dunleavy & Dede, 2014).

**Augmented reality as instructional application**

Regardless of the affordances mobile computing provides, sound pedagogy and instructional design remain primary determinants of success. In a project using GIS (geographic information system) software and GPS (global positioning system) to engage students in inquiry-based learning, Coulter and Polman (2004) concluded that focused curricular planning led to a more successful implementation than an “activity exposure” approach. Citing cognitive overload and unproductive mental effort as risks to learning in discovery-based learning environments that rely heavily on multimedia, Clark, Yates, Early, and Moulton (2010) support the use of guided training methods instead. Nevertheless, in a recent review of AR teaching and learning Dunleavy & Dede, (2014) contend that, “as a cognitive tool or pedagogical approach, AR aligns well with situated and constructivist learning theory as it positions the learner within a real-world physical and social context, while guiding, scaffolding and facilitating participatory and metacognitive learning process such as authentic inquiry …” (p.737).

**Research design**

The provenance for our methodology is informed by design-based research, where specific curriculum or instructional treatments are iteratively tested within real educational settings, (Parker et. al, 2011; Saye & Brush, 2017), and protocols developed by Graham Nuthall as part of the Project on Learning (Nuthall, 2000; 2005). Our work was designed as a unit of instruction within a locally-developed ambitious 5th grade curriculum (Kracjck & Blumenfeld, 2006, p. 326), entitled *My Place in Time and Space*, to investigate a forgotten local history via the site of a former African American school. Our research work is now in its fourth year and ongoing. Data for this paper is predominantly from 2014-2015.

**Instructional treatment**

Our perspectives for the design of instruction were guided by cognitive research that recognizes both the psychological and social aspects of learning (Anderson, 2009). Specifically, we pull from the How People Learn framework (Donovan & Bransford, 2005), situated learning theory (Anderson, Reder, & Simon, 1996; Brown, Collins, & Duguid, 1989), social constructivist learning theory (O’Donnell, 2012; von Glaserfeld, 1995; Vygotsky, 1978), the motivation to learn literature (Jones, 2009), and what Newmann and associates (1996) term Authentic Intellectual Work (AIW). As Shaffer and Resnick (1999) and Newmann et. al (1996) contend, authentic instruction occurs when learning is personally meaningful, relates to the outside world, provides opportunities to engage in substantive disciplinary conversations as a way to facilitate participation in a “knowledge creating culture,” and supports higher-order thinking skills that can be assessed appropriately. Sawyer (2008) also speaks to anchoring “authentic knowledge in its context of use” (p. 9). The idea of leveraging digital technologies as part of the process of “authentic instruction” to build theoretically-based and contextually aware scaffolds to support tightly bound disciplinary specific work is also informed by the literatures on the value and impact of historic sites/buildings to “read” and engage with the past and others across time and space (Baron & Dobbs, 2015), and the potential of problem based learning and more specifically technology enhanced problem based historical inquiry models (Brush and Saye, 2017). Our design leverages the power of place, the context of a problem and an explicit scaffold to embed authentic practices (historical inquiry) in a meaningful context, reduce the complexity of those practices and make them explicit (SCIM-C), and build upon prior knowledge (Edelson & Reiser, 2006, p. 336).

For this project, as detailed previously, a lone remaining building remains of the former 180-plus acre campus. This provided the basic context for our week-long fifth-grade unit. The unit itself would be framed around the question: “If this building could talk, what would it tell us about the people who were here 50 years ago and their experience?” In order to prepare students to explore this question on the site, the unit featured two separate in-class lessons that focused on harnessing requisite skills and digital literacies for the investigation. The fact that site was a segregated school was not revealed to students but rather reserved as part of the investigation and left to them to determine.

**Data collection and analysis**

The study included 14 fifth-grade classes, comprising 288 children within one local school division. Data collected included: field notes and video-recording of how the app was used in the classroom and on the field trip (this included student think-alouds as they used the app on the field trip); location data for each student group including
a log of the order of locations visited, time at each location, number of historical sources viewed, word count during analysis, and the number and range of sources students identified as significant; small group student interviews - post unit; student work samples generated as part of the unit - including detective journal and final products; teacher interviews - post unit.

All data was categorized by class and student teams so that interviews, data logs and student work could be aligned in order to trace the interactions and use of the app through to the final product. All interviews and video recordings were transcribed, indexed and underwent inductive content analysis to identify key themes around motivation to learn and explicit learnings and negotiations with the app (Sawyer, 2006, p. 13).

Findings

The project’s findings reflect the complexity of capturing fifth grade students’ capacities to think historically (both in-situ and post process) while tandemly assessing the role and impact of AR learning environments to support student historical inquiry and historical understanding. Our findings are considered within the context of the research question/perspectives guiding this work: Can we design an authentic problem based – place based – historical inquiry unit that utilizes augmented reality and an engagement-first approach to: 1) scaffold disciplined inquiry; 2) facilitate student understanding of change over time, and depth of understanding around school segregation; and 3) motivate and sustain students through the inquiry arc? The outcomes represent both generalized research findings across the 14 participating classes as well as a more specific look at the activities of one student group, documenting their experience and perspectives in-situ.

As part of the iterative design process, observations in the field and data log from our pilot test revealed the necessity to shift from the concept of unguided student exploration of the site (where students were tasked with simply exploring the site with the app), toward a guided exploration approach in order to sustain students within the inquiry arc. The guided exploration approach enabled many students to reach beyond the “wow” factor leveraged by the initial AR experience on site and sustainably engage with evidence through the process of inquiry on site. Once the higher level of guidance was in place, data log usage revealed a general increase in time spent with evidence and field observations noted a progression in sustained student engagement. The balance of guided exploration coupled with an appeal to student interest via augmented reality interaction on site proved a synergistic combination and contributive factor in sustaining inquiry with 5th graders.

With regards to artifacts, students identified photographs, newspaper clippings, and the daily class schedule as the most important evidence and these items were subsequently used most heavily in student work. Students placed less importance upon oral histories and transcripts and more text heavy sources and as such cited them less frequently within their final reports. Overall student final reports support the assertion that the AR unit contributed to student understanding of change over time and depth of historical understanding of school segregation further reflected in frequent empathetic stances invoked in student final reports.

Within many final reports, students make specific links to evidence and location as they develop a response to their question in a letter to the chief detective. What becomes clear in their final write up is that they directly connect to the evidence they investigate and also pay attention to the period and context from which the evidence emerged. Typically, and importantly, all groups were capable of identifying the space within which they were moving as previously being an African American School during the era of segregation that while prominent in the area was closed in 1966 as a result of integration.

Prior to developing their written reports, data collected also suggests that the built-in historical thinking scaffold (SCIM-C) within the app and the application of the engagement-first strategy helped students stay within the inquiry process and foster their understandings of the evidence in relation to the guiding question. In follow-up group interviews, students consistently voiced a positive interaction with SCIM-C and even reported they felt like historians working with evidence and interpreting the past. To this end, one student noted, “by taking the evidence through analysis, we weren’t just given knowledge and that made it more fun. We had to think about it more – and if all we did was read about it in a textbook, it’s really not going to stick in our heads.”

In analyzing sources using SCIM-C, students’ analysis and engagement supported Nuthall’s (2000) assertion regarding multiple stages of interaction and its relation to acquired understanding. In reflecting on the use of SCIM-C, one student added: “The questions helped a lot because it helped you see something that connects, so you want to go back and see and be like, oh, this connects with that.” Similarly, another student added: “I think the questions were really helpful because you didn’t just read about it on the other stuff. You read about what’s in the questions and you went back and actually looked at it. That kind of helps you learn more, like, if you go back and do it again and read it again and reread. That’s how I think the questions were really, really helpful.”

While it was possible to trace many team’s final written products directly to the evidence they noted in the virtual detective note books and stored in their virtual back packs it is important to note that the amount or level of actual written analysis stored in their virtual back packs was not necessarily always as detailed as we had
Analysis of response frequency within data logs showed that 94% of summarizing questions and 86% of contextualizing questions were attempted across all 14 fifth-grade classes involved; however, only 67% of monitoring questions and only 78% of corroborating questions were attempted (typed into the historical thinking scaffold within the app) in-situ. There was a clear discrepancy between how they talked through their analysis and what they actually recorded/wrote up on site. These results suggest that as scaffolding questions and phases of inquiry grow in complexity and sophistication, so too does the likelihood such phases will be abandoned or only minimally attempted when students are expected to record their ideas through typing.

However, for many groups, their final written reports clearly showed they had engaged with, paid attention to, and were now using and working across the sources to transform them into evidence to report that the site had been an African American school, while also providing details of these students experiences in the past at that site. What is also revealing is that even for those groups of students whose final written reports lacked the type of specific details that we hoped students would glean from working through the app our video recorded observations and accompanying prompted think-alouds of student teams reveals that even these students, through the use of the app could (1) identify and make reference to the buildings that were no longer there that formed part of the campus, and probably more significantly (2) verbally unpack individual sources in a systematic manner and (3) stay with and puzzle through evidence in order to draw thoughtful inferences. For example, videotaped observations of one group of students captures the moment that they struggle with the concept of “trades” as they investigated the virtual “Trades Building.” The evidence in the Trades Building included photographs and oral histories of students learning barbering and cosmetology. The group directly access, and refer to, specific evidence prior to making the claim that the site may have been an African American school. What becomes important here is that both students actively interact with their shared tablet as they look around the interior of the virtual building and both tap on the screen to access and look at specific evidence.

Student A: [holding tablet and pointing it across the street toward a construction site where the Trades Building once sat]. “So, we have to head toward the green one?” [At the same time student B touches the screen and hits the green titled trades building tab]. “Oh weird.” [Student B is trying to keep up with student A as he spins to his left. They are now inside the virtual building]

Student B: Are we inside? It looks like something at ... [both are touching the screen while student A is holding it and completes a 360-degree turn]

Student A: This was called the Trades. Is trades a word for Barbershop or something like that? Hey, is it an African American? That might prove some-

Student B: [Points at the screen] “Oh look, it’s a barbershop.”

Student A: Yes, it’s a barbershop. Why was it called trades? I wonder why it was called trades then? When it’s got barbering. [Students continue turning around the virtual building interior while looking through the tablet] let’s take a look at barbering … it might have been an African American school.

Observer: What makes you think that?

Student A: I think I might have found some evidence that proved it. There is like, here [points to photograph] there is barbering and it shows, like a picture of an African American, and there is like sound [oral history] ... So, I’m going to listen to that. [signals partner to get close as they play the oral history]

Additionally, video recorded observations of student teams also provide examples of students corroborating across sources to make evidence-based claims. Such corroboration reveals that even when teams were confused by some individual sources, they were capable of referring back to previous examined sources to sharpen their understandings of what they were now seeing and reading. Such corroboration, for a number of groups, occurred after visiting the third and final building – the Gym. For many groups, the issue of race and who went to the school only began to percolate as they began to specifically investigate the photographs of student activities, starting in the trades building and then more systematically in the virtual Gym. However, the fact that the photographs were black and white served, at the time, as an initial stumbling block to making the inference that the site was a segregated school. Interestingly, it was only when groups started to read a short newspaper article – located in the virtual Gym – about a performance of the glee club and band that they slowly came to infer and make the claim that this was an all African American school.
Student A: [Reading a newspaper report about a Glee Club performance aloud after telling the interviewer/observer that he thinks it might be an African American School, he quotes directly] “Negroes can excel in music and education” ... [looks at observer] [the school] is showing the public that Negroes can excel. Which means that to know this, they would have had to have Negroes at their school.

Observer: Do you see any other evidence that might suggest that it was an African American school?

Student A: Yes, because … um … [taps on tablet screen as both partners look at screen] in the other pictures [referring to the photograph in the virtual gym which is where they are now] and in some of the things inside of the Long Building, there was some talking about having African Americans here, or just colored people here at [the school]. And the pictures show that they are not white people. Although there are some white people in the picture. Like they look kind of white, but it is black and white pictures, so it is kind of hard to tell. Um … so, I think this is actually pretty important because all the other things like the newspaper [points at tablet screen], not the picture especially, but the newspaper [which he was reading earlier] is telling us that it is probably a negro or colored school. [begins one-fingered typing as part of analysis] so, errr … yep. [continues to type as partner watches]

What is also striking within the student’s discussion around the sources is that as he read the article and began to use it to make sense of what this site was in the past, he began to use words such as Negro and Colored that came directly from the source. His use of such language as part of his think-aloud illustrates the extent to which he was paying attention to the details of the sources while trying to use the sources to make a claim. Importantly, given the nature of the sources the students would interact with, prior discussions were had based on how and why such language would not be used today, but as part of an historical inquiry, and working with historical sources, one has to pay attention to the language and context of the period being investigated. Our taped observations begin to reveal how these students used the app to access sources and persist with an inquiry in-situ, while also beginning to get a sense of this site as a place with a history that was no longer visible. Their taped analysis also begins to reveal how even these young fifth grade students were able to begin use individual sources as evidence, before working across multiple, and sometimes confusing sources, to verbally corroborate and engage in a level of disciplinary analysis that, without being recorded, would have been lost and not evident within their final write up.

Discussion and conclusion

By specifically seeking to illuminate the impact of an augmented reality app to support student learning and specifically students’ abilities to think/express disciplinary thinking (both in-situ and post product), our findings reveal the complexity inherent in both designing instruction to support the inquiry arc and developing research protocols that make learning visible in terms of capturing the process of young students’ disciplinary thinking. These issues are amplified by navigating place-based history education, hidden and difficult histories, elementary students, and an engagement first approach to the acquisition of historical knowledge. What is clear and methodologically important is that if our analysis had purely focused on student work products in terms of their detective notebooks and final reports our team would have missed vital and positive insights gathered through participant videotaped observations. Videotaped observations that followed students through their entire field trip clearly illuminate how they used the app to engage with the sources in terms of summarizing, inferring, and corroborating across sources in order to verbally make evidenced based claims to answer the guiding question. While some of these claims were clearly evidenced in their final reports, this was not always the case and without capturing/recording their thinking through observations such important emerging disciplinary thinking skills would have remained invisible/ephemeral. This significant insight itself highlights the need for further refinement of the app in terms of shifting from having young students with little hands type their analysis to: (1) providing them with the option of voice dictation to capture their ideas and analysis; and (2) developing the type of real-time actionable and explicit strategic feedback to support students’ persistence and engagement as they unpack individual sources and look to corroborate across sources. Such feedback would replace the prompts from the researchers in the field that encouraged students to monitor their understanding as they progressed through the investigation.
Our findings illustrate the ability for AR, when combined with an explicit scaffold, to support young learners through an inquiry arc. The data collected illustrate a gap between the children’s ability to produce evidentiary warrant for their claims and our ability to capture that processing through quantitative data points. Only by carefully reviewing observation and interview transcripts was the process through which our young learners unpacked evidence and made inferences revealed. Much of the terseness in the students’ written responses was likely due to the simple fact that typing field notes on a tablet is not an easy task for a young person. However, despite hindrances like these students were capable of performing historical inquiry and creating their own evidence-based account of a local hidden history, as articulated through their think-alouds while on site (Sawyer, 2008, p. 6). The promise of this approach calls for further refinement of the experience and tools to better support historical inquiry and other problem-based approaches to learning.

References
Bain, R. B. (2005). “They thought the world was flat?”: Applying the principles of How People Learn in teaching high school history, In M. S. Donovan & J. Bransford (Eds.), How students learn: History in the classroom pp 179-213).
Johnson, A., Hicks, D., Ogle, T., Bowman, D., Cline, D. & Regan, E. (2017) “If this place could talk”: Using augmented reality to make the past visible. Social Education 81(2), 112-116.


Acknowledgements
This research was supported by the National Science Foundation under the Cyberlearning EXP Program, award IIS-1318977.
Toward a Taxonomy of Team Performance Visualization Tools

Zachari Swiecki, University of Wisconsin, Madison, swiecki@wisc.edu
David Williamson Shaffer, University of Wisconsin, Madison, dws@education.wisc.edu

Abstract: Research on teams has become increasingly important due in part to the status of collaborative problem solving as a vital 21st century skill. Much of this research has focused on factors that affect team processes and outcomes, such as the use of team performance visualization tools. Such tools are also valuable to researchers who make inferences from team data or educators who assess teams and plan interventions. This variety of users suggests that studying these tools requires a user-centered approach focusing on affordance relationships. In this paper, we use Epistemic Network Analysis to create a visual representation of the space of affordance relationships for extant team performance visualization tools. We use this space to compare tools, and to demonstrate empirically the dimensions along which they differ. These dimensions suggest a preliminary taxonomy of tools in terms of their affordance relationships for different users.

Introduction

Teams are collections of individuals who work together to learn, solve problems, make decisions, and design products. Recently, research on teams has become increasingly prevalent due to their ubiquity and growing importance in society. Moreover, collaborative problem solving—and collaboration more generally—has been recognized as a vital 21st century skill (Griffin & Care, 2014). Much of the extant research on teams has focused on factors that affect team performance, which may refer to either team outcomes or the processes teams take to achieve those outcomes. One such factor is how tool use by teams affects their performance. In particular, a growing body of research has focused on how visualizations of team performances affect team outcomes and processes (e.g., Fiore & Wiltshire, 2016; Janssen & Bodemer, 2013).

Team performance visualizations are not only important for teams, however. Given the complexity and volume of team interaction data that can be collected, visualizations also play an important role in how researchers investigate teams and how educators monitor and assess them. This variety of users, and their differing goals, suggests that those who study team performance visualizations should take a user-centered approach. Such an approach would consider both the properties of visualization tools and those who use them—in other words, it would consider the affordance relationships that exist between these tools and their users (Norman, 2013).

Several prior studies have investigated features of team performance visualization tools, such as awareness tools and dashboards (e.g., Bodily & Verbert, 2017; Janssen & Bodemer, 2013; Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). However, these studies are lacking either because they do not focus on team data specifically, they ignore certain user groups, they ignore large classes of tools, or they do not consider how the affordances of tools relate to one another. To address this gap, we conducted an empirical investigation of extant team performance visualization tools and derived a preliminary taxonomy in terms of their affordance relationships with three user groups: teams, researchers, and educators. Such a taxonomy should help to better distinguish tools and provide guidance to different users. To analyze and visualize our data, we used Epistemic Network Analysis, or ENA, (Shaffer, Collier, & Ruis, 2016) a method for analyzing the connections between features of interest in data, and thus a powerful technique for investigating the collections of affordances that exist in tools.

Theory

It is widely recognized that the tools we use affect how we think and act (Hutchins, 1995; Vygotsky, 1978). Several fields, including the learning sciences, social psychology, and organizational psychology, have investigated tools in relation to teams. Much of this work has focused on the effects of different visualization tools on team performance. For example, those who study Computer Supported Collaborative Work have investigated how shared representations, such as digital whiteboards, affect team performance (Fiore & Wiltshire, 2016). And in Computer Supported Collaborative Learning, researchers have studied how different visualization tools affect communication (Munneke, Andriessen, Kanselaar, & Kirschner, 2007), the construction of joint problem spaces (Roschelle & Teasley, 1995), and participation (Janssen & Bodemer, 2013).

Team performance visualization tools—and tools more generally—affect teams through what Norman (2013) calls affordances, or particular relationships between the properties of tools and the capabilities of their users. Several authors have investigated the affordance relationships that exist between team performance
visualization tools and teams. For example, Janssen and Boedmer (2013) argue that cognitive group awareness tools, which provide information about the distribution of team knowledge, help to coordinate social and cognitive activity by facilitating communication and the development of shared mental models. Similarly, social group awareness tools, which provide information about the activities of team members such as communication levels or contributions to tasks, help teams to coordinate action and information in situations where they cannot directly see one another’s behavior. These awareness tools may also allow team members to identify and directly compare their performance to that of their teammates, creating opportunities for social comparison. Other tools may occlude individual contributions, instead showing metrics of team performance only at the team level (Kimmerle & Cress, 2008; Kimmerle, Cress, & Hesse, 2007).

While it is clear that tools affect how teams perform, they also affect how researchers measure team performance. To understand and make claims about teams, researchers design assessments to collect evidence that warrant those claims (Mislevy, 1996). However, teams are dynamic and complex systems. They have multi-level structure, individuals interact with one another and with tools, and these interactions change over time. The combination of these features means that evidence elicited from teams is both vast and complex. For example, many researchers study teams who work together in virtual environments. Such environments log data on interactions within the environment, including mouse clicks, verbal or chat communications, and product submissions. The complexity of these data has led many researchers to use visualization tools in order to make meaning from the evidence they collect. For example, researchers have used a variety of network tools to visualize both the structure of team interactions and the semantic content of those interactions (e.g., Dawson, Tan, & McWilliam, 2011; Sha, Teplovs, & van Adst, 2010). Temporal visualization tools, such as CORDTRA, have been used to analyze the sequence and simultaneity of team interactions (Hmelo-Silver, Jordan, Liu, & Chernobilsky, 2011). And dependency visualizations, such as those derived from state transitions, Markov models, and process models, have been used to investigate the likelihood that team actions follow or depend on one another (Reimann, P., Frerejean, J., & Thompson, K., 2009).

Furthermore, given the volume and complexity of team data, educators have come to rely on tools that convey information about teams succinctly in real-time. Consequently, learning analytic dashboards have been designed to provide educators with information about team activity and performance in condensed views (Verbert et al., 2013). Many of these dashboards, such as the Process Tab (Shaffer, 2017), which uses network diagrams to represent the connections teams make between concepts as they collaborate, allow educators to access the data underlying the visualizations to see more detailed information on demand.

Each of the properties described above—knowledge, activity, member identification, structure, semantics, temporality, dependence, and details—are features of team performance visualization tools that afford different actions for different users. Teams use these affordances to facilitate communication and compare their performance; researchers use them to test hypotheses and build theories; and educators use them to assess students or plan interventions. Importantly, the affordances of any given tool do not exist in isolation. For example, concept maps, one kind of cognitive group awareness tool, represent information about team knowledge and information about the structure of that knowledge—that is, how knowledge components relate to one another. In addition, dashboards typically contain a variety of visualizations of team performance. One dashboard, Collaid (Martinez Maldonado, Kay, Yacef, & Schwendimann, 2012), contains at least five visualizations, including radar graphs, pie charts, and time series, which simultaneously provide information about how team activities change over time. In other words, team performance visualization tools are interactive systems in which different affordances relate to and affect one another. Thus, understanding and describing the affordance relationships between different tools and their users requires a method of analysis focused on how affordances connect to one another. One such method is Epistemic Network Analysis, or ENA (Shaffer, Collier, & Ruis, 2016).

ENA measures connections among relevant features in data and represents those connections as networks. ENA was developed to model cognitive networks, or networks that represent associations between elements of complex thinking. This method has been used to study the patterns of connections that teams of learners make as they solve problems in simulations of professional practice (e.g., Shaffer et al., 2016). Thus, ENA itself is a team performance visualization tool. However, ENA is also a general method for investigating data where complex patterns of relationships are thought to exist. In this study, we use ENA to create a visual representation of the space of affordances for extant team performance visualization tools. Using this space, we compare tools, and also demonstrate empirically the dimensions along which they differ.

Several prior studies have investigated the affordances of team visualization tools for different user groups. However, each case has important limitations. For example, Verbert and colleagues (2013) conducted a review of over 20 learning analytics dashboards. These authors classified each dashboard according features such as target user (teacher or student) and data source (social interaction, time spent, and so on). This review accounted
for two kinds of users; however, it was limited to one particular kind of visualization tool and did not specifically target dashboards that represent team performance data.

Taking a team focus, Janssen and Bodemer (2013) reviewed the literature on group awareness tools. These authors identified the kind of data represented by each tool, how the data was collected, and characteristics of the visualizations used. This work explicitly looked at tools designed for teams as the target user, and it considered several affordances of these tools for teams. However, other aspects of the visualizations considered tended to focus only on surface level properties and not the interactions they afforded. Moreover, this review did not consider tools designed for researchers or educators.

Each study above contributes toward understanding the affordance relationships between team performance visualization tools and their users. However, they either ignore certain users, ignore large classes of tools, or do not focus on teams. Work discussing team performance visualization tools for researches is also deficient in that it often only implicitly demonstrates their affordances. That is, this work shows what these tools can do in terms of providing insights into team performance or building theory, but they do not explicitly discuss the properties of the visualizations that make these insights possible. And while some work is more explicit in this regard (e.g., Siebert-Evenstone et al., 2016), it is rare that authors discuss a variety of these visualizations together in one place.

In this paper, we address the limitations of prior work by conducting an empirical investigation of the affordances of extant team performance visualization tools. In particular, we examine the literature on these tools and ask the following two research questions: (1) what are the affordances of extant tools? And, (2) how do these affordances differ across user groups? This analysis provides some first steps toward a taxonomy of tools that should guide future work and help users select appropriate tools given their goals.

Methods

Data
To address the research questions above, we first collected a representative sample of papers that described team performance visualization tools and their affordances. To collect this sample, we consulted a panel of three domain experts who have extensive experience in measuring and visualizing team performance. These experts suggested specific papers in which team performance visualization tools were described, as well as topics for a literature search. Using these papers, their citations, and literature search topics, we collected a sample of 41 papers and book chapters, which were reviewed and approved by the expert panel. From this sample, we identified 48 team performance visualization tools for this study.

Coding
Two raters classified the tools described in each paper into one of six types: (1) social group awareness tools, (2) cognitive group awareness tools, (3) dashboards, (4) networks, (5) dependency visualizations, and (6) temporal visualizations. The distribution of tools by type is shown in table 1.

Table 1: Counts of team performance visualization tools by type

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Group Awareness</td>
<td>17</td>
</tr>
<tr>
<td>Cognitive Group Awareness</td>
<td>11</td>
</tr>
<tr>
<td>Dashboards</td>
<td>11</td>
</tr>
<tr>
<td>Networks</td>
<td>4</td>
</tr>
<tr>
<td>Dependency Visualizations</td>
<td>3</td>
</tr>
<tr>
<td>Temporal Visualizations</td>
<td>2</td>
</tr>
</tbody>
</table>

Next, the two raters coded each tool for their intended user and affordances. The user categories included four codes: (1) team, (2) educator, (3) team and educator, and (4) researcher. Tools coded for the first three categories were described in papers that explicitly mentioned the end user of the tool. Tools coded for the researcher category either did not mention the end user of the tool, or only reported on the analysis of team performance data using a given tool. In total, 31 tools were coded for teams as the intended user, 3 for educators, 5 for teams and educators, and 9 for researchers. We developed the visualization affordance codes from a grounded analysis (Glaser & Strauss, 1967) of the data. We define these codes in table 2.
Table 2: Visualization affordance codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>Represents information about the activity or participation of the team, e.g., how much they are communicating or contributing to the task.</td>
<td>Collaid dashboard: Martinez-Maldano et al., 2012</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Represents information about the knowledge, opinions, or understandings held by the team.</td>
<td>Concept maps: Engelmann &amp; Hesse 2011</td>
</tr>
<tr>
<td>Structure</td>
<td>Represents the structure of teams, their interactions, or their knowledge, e.g., subgroups, communication patterns, or connections between concepts.</td>
<td>Social network graphs: Dawson &amp; McWilliam, 2011</td>
</tr>
<tr>
<td>Semantics</td>
<td>Represents the content of team interactions or knowledge—rather than, or in addition to—quantity or structure.</td>
<td>KSV: Sha et al., 2010.</td>
</tr>
<tr>
<td>Dependence</td>
<td>Represents information about how team actions depend on other team actions, e.g., the likelihood that one action follows another.</td>
<td>Dependency graphs: Reimann et al., 2009.</td>
</tr>
<tr>
<td>Temporality</td>
<td>Represents information about the simultaneity, sequence, or change of team actions or knowledge over time.</td>
<td>CORDTRA: Hmelo-Silver et al., 2011.</td>
</tr>
<tr>
<td>Member Identification</td>
<td>Contributions from each individual team member are identifiable within a single visualization in the tool.</td>
<td>Social awareness tool: Kimmerle et al., 2007.</td>
</tr>
<tr>
<td>Details</td>
<td>Allows users to access the data underlying the visualization.</td>
<td>Process Tab: Shaffer, 2017.</td>
</tr>
</tbody>
</table>

Two raters coded each visualization tool for the type, user, and affordance codes described above. Given the relatively small sample size, both raters coded the entire data set together, resolving differences as they coded, and arriving at a single decision for each code. Thus, the kappa statistic for this coding was 1.

Analysis
After coding, we analyzed the data using ENA. ENA measures connections among codes in data and represents them as networks. In this study, we used ENA to construct network models for each of the 48 team performance visualization tools. Here, the units of analysis are the tools, the network nodes correspond to the affordance codes in table 2, and a network for a given tool represents its collection of affordances.

The process of creating ENA models is described in more detail elsewhere (Shaffer et al., 2016), but in brief, ENA creates adjacency matrices for each unit of analysis that quantify the co-occurrences of codes within the unit. Next, the adjacency matrices are normalized and represented as vectors in a high dimensional space, where each dimension corresponds to a co-occurrence of codes. A dimensional reduction via singular value decomposition is then performed to project the vector for each unit of analysis into a lower dimensional space that maximizes the variance accounted for in the data. Finally, the nodes of the networks are placed in the space using an optimization algorithm such that the center of mass for a given unit’s network closely corresponds to that unit’s location in the lower dimensional space. Importantly, these nodes placements are fixed, meaning that the nodes of each network are in the same place for all units in the analysis. This fixed set of node positions allows for meaningful comparisons between units in terms of their connection patterns and allows us meaningfully interpret the dimensions of the space.

The final result is two coordinated representations for each unit of analysis: (1) a plotted point, which represents the location of that unit’s network in the low-dimensional space, and (2) a weighted network graph. Because the location of any plotted point corresponds to the center of mass of its corresponding network, we can use the weighting of the network to explain that point’s location. Thus, plotted points located on the right hand side of the space will have networks whose strongest, or most heavily weighted, connections appear on the right. Similarly, plotted points located on the left will have networks whose strongest connections appear on the left, and so on. Using ENA, we can also group plotted points by metadata and calculate the mean position in the space for a group. This mean can be plotted in the same metric space and has a corresponding network graph which represents the average connections between codes for that group. In this analysis, we grouped the visualization tools in two ways: by type of tool (table 1) and by user. In both cases, to make comparisons between groups we used their mean position in the space, their mean network diagrams, and network difference graphs, which subtract the edge weights of two networks to show the strongest connections in each group.
Results
In this study, we used ENA to investigate (1) the affordances of extant team performance visualization tools and (2) how those affordances differ across user groups. With regard to the first, in the figures below we see the results of the ENA analysis for the six types of tools described above. Fig 1 shows the mean plotted points (colored squares) for each type of tool, as well as the overall mean point, in ENA space. In the figure, we have also included the network for the overall mean, which helps to interpret the dimensions of the space. For the mean network in figure 1, the thickness and saturation of a given connection reflects the average number of tools coded for both codes defining the connection.

Fig 1. Mean plotted points and overall mean network for the six tool types. Legend on the right.

Because of the relationship between network representations and plotted points in ENA, we can use this network layout to interpret the dimensions of the space, and in turn, the meaning of each point’s location. Thus, the Y dimension separates visualization tools in terms of their focus on more complex data transformations, such as showing temporality, dependence, or structure (top), versus less complex transformations like showing individual contributions or underlying data (bottom). On the other hand, the X dimension separates visualization tools in terms of their focus on social information about each team member, such as their activity levels (left), versus more cognitive information, such as the knowledge held by the team (right).

The interpretation of the dimensions above helps to explain the location of tools in the space, as well as the differences between them. For example, in the figure, we see that the mean location of social and cognitive awareness tools in the space is quite different. To highlight this difference, in figure 2 we constructed a difference graph showing which connections, on average, are stronger in one kind of tool versus the other. In particular, this graph shows that cognitive awareness tools (red) are toward the right and lower down in the space due to their focus on representing team knowledge, underlying data, and the contributions of all members on the team. Contrastingly, social awareness tools (blue) are toward the left and higher up due to their focus on representing the social activity of each team member and how that activity changes over time.

Fig 2. Network difference graph for social (blue) and cognitive (red) awareness tools.
Page limitations do not allow us to include the network diagrams like those in figures 2 for each type of tool; however, the defining affordances of each tool type can be inferred from the location of their mean plotted points and the interpretation of dimensions described above. For a more detailed account, we examined the separate networks for each type of tool and found that network visualizations (purple, figure 1) are defined by more complex representations of the structure and semantics of team activities and knowledge. Similarly, dependency visualizations (in orange) are defined by the representation of those same features, with the addition of information about how team activities or knowledge components depend on one another. Temporal visualizations are defined by representations of temporality and the semantic content of team activity and knowledge. Finally, dashboards are defined by representations of how the activity of team members change over time.

With regard to our second research question, figure 4 below shows the results of the ENA analysis of visualization tools grouped by their intended user. Thus, the colored squares in this figure correspond to the mean plotted point positions of tools grouped by user, whereas in figure 1 above, the colored squares correspond to the mean plotted point positions of the tools grouped by visualization type. Here again, we include the overall mean network diagram to aid in interpretation. Because only the groupings of tools have changed, the interpretation of the dimensions remains the same as above.

Figure 3 shows that, on average, visualization tools designed for teams (in blue), focus on representing information about the activity and knowledge of each individual on the team, as well as the data underlying the visualizations. On the other hand, tools designed for use by researchers (in red) focus on representing more complex information such as temporality, structure, and dependence.

Tools designed for educators or teams and educators (green and purple respectively) are located below and to the left of researcher tools, meaning that they focus more on the activity of team members and underlying data, rather than more complex team performance data. However, their location on the Y axis is higher than tools designed for teams, meaning that they focus relatively more on complex data than team tools, on average.

**Discussion**

The ENA analysis above suggests a preliminary taxonomy of team performance visualization tools in terms of their affordances. The dimensions of the ENA space distinguish between those tools that focus on representations of social versus cognitive information on the X axis, and those tools that focus on more or less complex transformations of team performance data on the Y axis. These dimensions define four broad categories with which we can classify the different tools: complex/cognitive, complex/social, simple/social, simple/cognitive. Thus on average, network, dependency, and temporal visualizations fall into the complex/cognitive category; dashboards fall into the complex/social category; social awareness tools into the simple/social category (though just barely); and cognitive awareness tools into the simple/cognitive category. These same categories also apply to the tools when grouped by user: researcher tools tend to be complex/cognitive; educator tools, complex/social; educator and team tools, complex/social; and team tools, simple/social.

The taxonomy described above has at least two important implications, one practical and one theoretical. First, it allows potential users to make clear distinctions between tools in terms of their affordances. This
taxonomy could help users to distinguish between broad types of visualization tools such as temporal versus network tools or to make more fine-grained comparisons between individual tools. These distinctions may also help users understand which tools complement one another, and guide their decisions about which tools to use given their goals. For example, researchers interested in visualizing complex cognitive information about teams over time may find temporal visualizations useful. However, if they are also interested in the structure of team cognition, or how team cognitive components depend on one another, then they may want to supplement their analyses with network or dependency tools.

Second, this taxonomy suggests potential areas of further inquiry and visualization tool development. For example, the discrepancy between tools designed for researchers and those designed for teams suggests that more complex and cognitively focused tools have yet to be developed for teams. A potential explanation for the lack of more complex tools for teams may have to do with the constraints associated with them. Just as tools have affordances that make certain interactions possible, their constraints place limits on those interactions. The researcher visualization tools are constrained in the sense that they may require more knowledge or expertise to understand, and more time to interpret. Teams that lack this knowledge and time may not benefit from such tools unless they are carefully designed for their contexts. For now, the effect of such tools on team performance remains an open question.

This study has several obvious limitations. First, our data selection procedure was limited in that it relied on the suggestions of a small panel of experts. Although these selections were supplemented by citation and literature searches, a more systematic search of the literature would likely have added to our sample. Future work will explore more systematic data selection methods. Second, in the majority of cases, our coding process was limited to investigating descriptions or images of the visualization tools rather than the tools themselves. Thus, it is possible that some important features were missed if they were not clearly described in their write-ups. Finally, the sample sizes for some types of visualization tools, such as temporal tools, were relatively small. The normalization feature of ENA should mitigate such differences in sample size; however, future work will target literature searches around tools that were less represented here. Despite these limitations, this work provides a preliminary taxonomy of team performance visualization tools based on an empirical analysis. Investigating more tools will test the stability of the ENA model and taxonomy presented here, and should improve its utility for teams, educators, and researchers.

References


**Acknowledgments**

This work was funded in part by the National Science Foundation (DRL-0918409, DRL-0946372, DRL-1247262, DRL-1418288, DRL-1661036, DRL-1713110, DUE-0919347, DUE-1225885, EEC-1232656, EEC-1340402, REC-0347000), the MacArthur Foundation, the Spencer Foundation, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.
Is Scrolling Disrupting While Reading?

Katherine Brady, Sun Joo Cho, Gayathri Narasimham, Douglas Fisher, and Amanda Goodwin
katherine.a.brady@vanderbilt.edu, sj.cho@vanderbilt.edu, gayathri.narasimham@vanderbilt.edu,
douglas.h.fisher@vanderbilt.edu, amanda.goodwin@vanderbilt.edu
Vanderbilt University

Abstract: The relationship between scrolling and comprehension has been studied a few times by comparing readers who read in an environment with scrolling to those who read in an environment without scrolling. However, the amount of scrolling done by subjects who had the option to scroll has not been linked to comprehension before. In this paper, we investigate the relationship between how often participants scroll and comprehension. We find that participants who have no preference between paper and digital reading tools do worse when they scroll, but there is no relationship between scrolling and comprehension for other students.

Introduction
As we enter the digital era, it is important to understand how design choices in digital reading environments affect comprehension. In this study, we examine how scrolling, the most popular method of reading long text passages on digital devices, correlates with comprehension. Previous research has found some evidence suggesting that scrolling is worse for comprehension than paging. Research comparing reading on paper to reading on digital devices found that there is no significant difference when the stimuli fits on one screen, but students do worse on digital devices when the stimuli takes up more than one screen (Kim & Huynh, 2008; Singer & Alexander, 2017). Other research suggests comprehension is better for digital readers using a paging interface than those using a scrolling interface (Piolat, Roussey, & Thunin, 1997; Sanchez & Wiley, 2009). These findings are in line with research that shows we use spatial markers to remember what we have read (Baccino, 1994; Rothkopf, 1971; Weger & Inhoff, 2007). When these spatial markers move, as happens while scrolling, readers cannot remember what they read as well, or find previously read passages as quickly.

This study is the first to examine how scrolling frequency is linked to comprehension. If scrolling hurts comprehension, then learners who scroll frequently while reading should perform worse on an open book quiz than those who move the text less often. We recorded the screens of 381 5th to 8th graders in the United States while they read a passage on a laptop and then took an open book test. Scrolling was measured by pulling out a frame from each participant’s recording every 0.1 seconds. If the text moved between two frames, that transition was counted as a scroll. We found that students who had no preference between digital and paper reading performed worse when they scrolled frequently, but scrolling had no effect on the learning of other students. We hypothesize that learners who have a medium preference have developed that preference through exposure to both mediums, and are more likely to have strategies for scrolling.

Background
In the most recent synthesis of comparisons between digital and paper reading, Singer and Alexander (2017) conducted a survey of 15 studies occurring between 2001 and 2017 that compared digital and paper reading for texts of different lengths. Seven of these studies used a stimulus which was 500 words or more, and eight used a text which was under 500 words for their stimulus. Singer and Alexander found that the eight studies which used shorter texts either noted there was no difference in comprehension between digital and paper reading, or found that digital reading was better for comprehension. In contrast, six of the seven papers which used longer texts found that students comprehended better in the paper condition. The one exception exclusively studied second language learners and noted that the dictionary lookup functionality in the digital condition was popular. Singer and Alexander hypothesized that the reason digital readers consistently did worse than paper readers on comprehension tests for long passages but not short passages is that the most popular method of navigating long text on the computer screen, scrolling, negatively affects reading comprehension. This hypothesis is supported by a 1997 study that compared scrolling and paging as methods of interacting with digital text. The 1997 study found that the summaries written by undergraduates who paged through a digital document were more coherent than those assigned to navigate the same document with scrolling (Piolat et al., 1997). Part of these results may be explained by the poor user experience of scrolling in many applications in 1997. A study of 10 computer scientists in 1997 found that scrolling mechanisms were very slow and users reported being annoyed that to scroll they had to click on a small scroll bar and that the display did not refresh fast enough to show them their position while they were scrolling (O’Hara & Sellen, 1997).
More recent studies suggest there are some negative effects from scrolling, but they are not as strong as those Piolat et al. found in 1997. In 2007, a study of Swedish participants found that those who had to use a mouse pointer to scroll while reading reported more stress and mental fatigue than those who had to use the keyboard arrows to page through the document (Wästlund, 2007). However, there was no difference in how those who navigated by scrolling and those who navigated by paging performed on comprehension exams after reading. In 2009, a study of American undergraduates found that among students who were either less familiar with the subject matter of the stimuli or less familiar with reading webpages, comprehension test scores were higher if they clicked through digital pages than if they scrolled while reading (Sanchez & Wiley, 2009). For students familiar with the stimuli or familiar with reading webpages, there was no difference between navigating by scrolling and navigating by paging. A 2011 study of Polish adults found that participants who paged through paragraphs and those who scrolled through the document performed equally in a recall test, but those assigned to the scrolling condition took longer to read the document (Kłyszejko et al., 2011).

Together these studies suggest the impact of scrolling on reading processes and comprehension may be decreasing as the user experience improves and familiarity with scrolling increases. The question we explore in this study is whether scrolling in today’s digital age would impact reading or whether perhaps familiarity with this process has reached a point where it no longer influences comprehension and memory. Additionally, each of the studies above compared the occurrence of scrolling to a different behavior (i.e., page turning or clicking through pages). This assumes all scrolling behaviors are similar and have a similar impact on reading. Because readers vary in how they scroll, our study examines the quantity and directionality of scrolling as it relates to comprehension.

One of the reasons scrolling may affect reading, is that semantic memory of text is linked to spatial memory. This was discovered by studies like Rothkopf’s 1971 experiment, which found that substantive memory of sub-passages in a document was directly related to incidental information about the spatial location of the sub-passage (i.e. which page it was on; Rothkopf, 1971). More recent studies report similar findings, suggesting the arrival of digital age has not changed readers’ reliance on spatial mapping of content. A study of French speakers in 1994 found that after reading a sentence on a computer, subjects could accurately point to where on the screen each word in the sentence had been displayed (Baccino, 1994). In 2007, Wegner and Inhoff used an eye tracker during a reading session (Weger & Inhoff, 2007). They asked readers to look at words they had just read and found that subjects’ eyes immediately found the words on the screen without needing to re-read the passage. These experiments show that without the support of spatial mapping, the cognitive load on readers is increased. Since text in a scrolled document doesn’t have a fixed position like text on a page, readers who scroll more may perform worse on comprehension tests because they are forced to remember content rather than map content. In contrast, readers who scroll fewer times, like those who only scroll when they must access additional text, may have a lighter cognitive load because they may still be able to map content onto space.

The question remains, though, as to whether readers’ increased familiarity and quantity of digital reading experiences have changed how they spatially map content in digital text. For example, a digital reader may have developed compensatory strategies for dealing with the fluid nature of digital texts. Accustomed to scrolling, a reader may map content onto other content, which also moves, like an image or heading providing a constant location in relation to the other content in the text even as scrolling occurs. This may provide a similar decrease in cognitive load as mapping content on space within a physical page. Additionally, it may mitigate the negative impacts of scrolling found in earlier studies of digital reading.

All this assumes, though, increased familiarity with digital reading behaviors like scrolling, which have gained a lot of popularity since the late 90s. A study of 15 German university students in 2010 found that most participants preferred to read a small area of the screen and scroll the text they were reading into that area to reading the whole screen and then scrolling to the text (Buscher, 2010). When questioned about their web reading behavior, experienced users have reported that they use scrolling to move distracting ads off their screen while reading webpages (Hillesund, 2010). Perhaps the biggest proof that user interfaces for digital text are improving and gaining popularity is that more users are opting to read long forms of text on digital devices instead of printing documents to find which contained relevant information for a search query (Kelly & Belkin, 2001). Researchers

In this paper, we wish to investigate scrolling in today’s digital era. We examine variability of scrolling behaviors and whether scrolling impacts reading comprehension or is an equivalent form of navigation. We conjecture that if scrolling creates more of a cognitive load, then learners who scroll more will perform worse on comprehension and recall tests than learners who scroll the text a whole screen’s worth only when they must.

There have been a few studies which counted the number of times a participant scrolled, but to the best of our knowledge, none have been conducted on the scale described here and none have tied the total amount of scrolling to performance. For example, in 2001, six American academics were recorded searching through 561 documents to find which contained relevant information for a search query (Kelly & Belkin, 2001). Researchers
recorded how often users scrolled in each document by recording the number of times they clicked on the scroll bar. They found that there was no difference between the number of times users scrolled when viewing relevant documents and the number of times they scrolled while viewing irrelevant documents. In 2016, Freund et al. recorded the screens of 41 Canadian participants and counted the number of times they scrolled to the bottom to check the length of the article they were reading (Freund et al., 2016). Readers checked the length of the document they were reading more often when the stimulus was a typical webpage with distracting advertisements and pictures than when it was stripped of extraneous content. This suggests scrolling behaviors differ, but does not link such behaviors to comprehension, which is the goal of the current study.

Method
We recorded the screens of middle schoolers while they read a website. We then used a computer program to calculate when they scrolled, and checked the results of the program against our own intuition by looking at visualizations of students at the far ends of the spectrum and summary statistics such as the average length of a student’s scroll. Once scrolling was calculated for all participants, we linked it to comprehension using a statistical model which took other factors such as prior knowledge into account.

Experimental setup
Our participants were 381 fifth to eighth graders (N=87 fifth graders, 78 sixth graders, 83 seventh graders, and 132 eighth graders) who were learning in the classrooms of 11 teachers across 3 schools in an urban district in the Southeastern United States. Eighth graders were oversampled because the stimulus was a passage designed for 8th graders by the National Assessment of Educational Progress (NAEP). We received demographic data on 369 of the participants from the school board. Two hundred and nine (56%) of the students were female. 165 were White, 157 were Black, 34 were Hispanic, 12 were Asian and 1 was an American Indian. According to the 2010 census, our sample had about 12% more Blacks and less Whites than the town’s distribution, but otherwise matched city-wide demographics. There were 87 students noted as economically disadvantaged and 66 spoke a language other than English at home.

All students did a mix of digital and paper reading as part of their regular classwork according to their teachers, though no classroom used one-to-one digital devices in class. All digital reading was recorded by iMotions (iMotions A/S, 2016). The recording contained an estimate of where the participant was looking aligned with the digital content being viewed, but we only used the screen recording for this study. For the content, we divided an NAEP passage (https://www.nationsreportcard.gov/reading_2011/testyourself_g8_passage_ann.aspx) into two parts with the dividing point chosen due to the natural end of a section. Therefore, the first part had 434 words and the second part had 674 words. Each participant read one part on paper and one part on a laptop with the order of the reading randomly assigned. Students were interviewed individually in a quiet classroom in their school. We measured student familiarity with the topic and preferences regarding digital and paper reading tools using a pre-test. After reading both sections, they took an open book post-test in which most of the questions referred to specific sentences or paragraphs in the stimuli.

When students started the digital reading, the researcher showed them how to use the highlighting, dictionary lookup and touch screen tools which were available in the digital environment. We did this because these tools change enough between devices and platforms that it was unlikely all students were already familiar with the tools we used, though their mechanisms were the same as similar tools found in PDF viewers.

Calculating scrolling
We developed a computer program to analyze the screen recording video. Note that both passages were too long to fit on a single screen, and therefore participants were required to scroll down to see the full content. To determine number of scrolls, the program pulled out frames from the video every 0.1 seconds and used image processing to figure out where each line of text was relative to the top and bottom of the screen as well as which lines were present in a frame. This information was used to calculate how far down the page a participant had scrolled and whether they had zoomed into the screen. To calculate scrolling, we listed the frames in the order they appeared in the video. If the text moved more than half a line’s height between one frame and the next, that counted as a scroll. We used a cutoff above zero because the image processing was occasionally off by a few pixels. This method worked for 378 of the 381 students we gathered data on. The screen capture program crashed for two of the students, and for one student the image processing algorithm was unable to identify the text in the frames due to the student highlighting all the text on the screen. The code for this program can be found at https://github.com/kbrady/eye-tracker-pipeline.

In addition to the total number of scrolls, we also calculated how many scrolls went up versus down on the page. We theorized that scrolling up to possibly review the document would have a different relationship with
comprehension than scrolling down. One hundred and forty-one students (37%) never scrolled up after the experimenter showed them how to scroll and 241 (64%) had less than five frame transitions in which they scrolled up. For perspective, the medium number of frame transitions in which a student scrolled down after the instructor gave them control of the computer was 14. This meant that for most participants, the number of timeframes they scrolled down and the total number in which they scrolled ended up either being the same or very close.

Since the laptop used to display the text was a touch screen, a lot of participants opted to zoom in to see the text or picture better (109 of the 378 students we got data for zoomed in at least once). If the zoom level on two consecutive frames was not the same, we did not count that pair as a scroll even if the text moved vertically. Due to the sensitivity in multiple directions of touch screen interface, it was very easy for students to accidentally scroll while zooming. For this reason, we felt it was best not to count text movement during a zoom as a scroll.

Visualizing scrolling
We checked that the scrolling measured by the computer program matched our own intuition of how much users scrolled while reading by visualizing each participant’s scrolling. This allowed us to see the scroll positions detected by the computer, compare them to the screen recording, and quickly check whether the ranking of participants by number of scrolls made sense. An example visualization is shown in Figure 1. This image shows the median student measured by number of scrolls while reading the second part. The image on the left shows the vertical position of the screen plotted against the time since the session began. The three outtakes on the right-side show what the student’s screen looked like during three sequences when it was stable. The top outtake shows what the student saw from 48 seconds after the session started to 2 minutes and 25 seconds. The middle panel shows what the student saw from 2 minutes and 26 seconds in to 2 minutes and 52 seconds. This outtake corresponds to the shaded area above the 2:47 tick in the graph on the left. The final outtake shows what was on the student’s screen from 2 minutes and 53 seconds after the session started to 3 minutes and 48 seconds after the session started.

The x-axis in the graph is used for time. In general, students did not zoom in enough to cut off the horizontal sides of the screen, so they could see the full horizontal width of the text (as shown in the outtakes to the right in Figure 1). Thus, the shaded area in the graph only represents the vertical position of the screen. We overlaid the plot on a stitched together picture of the whole document so we could see which parts of the document were visible to the student at each time point during their reading session.

Figure 2 shows the same visualization applied to the students who scrolled the most and the least during the second part. The frequent movement of the blue band in 2.a shows that the student scrolled quite a bit, while the stability of the blue band in 2.b shows that the student only scrolled in one short burst a little over 7 minutes into the session. Visualizing scrolling in this way also allowed us to see certain macro behaviors. For example, the thin blue downward spikes in the first minute of Figure 1 and Figure 2.a show that the screen was moved down and up very quickly shortly after the page loaded. This was the period in which the researcher showed the participant how to use the tools in the digital environment, including scrolling and zooming on the touch screen.
Another macro behavior is reviewing the document, which is shown after 3:30 in Figure 2.a. We believe training the computer program to recognize and classify these behaviors would allow us to gain further insights into how students scroll while learning.

![Figure 2. Examples of students scrolling while reading the second section.](image)

We also wanted to check whether students who scrolled more, scrolled shorter lengths than their peers. As shown in Figure 2, students on the far ends of the distribution seemed to support this theory, but we did not have time to look through the visualizations for all students. Therefore, we plotted the average length of a downward scroll for each condition against the number of times a student scrolled. As shown in Figure 3, there was a relationship between the number of times a reader scrolled and the average length of a scroll, with those who scrolled less often making longer scrolls.

![Figure 3. Average length of a downward scroll in pixels plotted against the number of times a student scrolled while reading the first section.](image)
Statistical models
We linked scrolling to post-test results (i.e., comprehension) using generalized linear mixed models. Each model was fit using the Laplace approximation for maximum likelihood. The model took pre-test scores, grade level, section read digitally, reading modality preferences and ethnicity into account as covariates. In addition to directly testing whether the number of times a subject scrolled had a relationship with comprehension, we considered the effects of holding these variables controlled. There were eight students who did not have a complete post-test and therefore had to be left out of the model. With the three students whose video recordings could not be processed, this limited the final model to 370 participants.

Results
Among students who had no preference between reading digitally versus on paper, we found a significant negative relationship between scrolling down and comprehension, controlling for the other covariates in the model. This group represented 39% of the subjects and was larger than either the set of students who said they preferred reading on paper (24%) or the set who said they preferred to read on digital devices (37%). The students without a preference between digital and paper were disproportionately likely to be non-white, but otherwise seemed demographically equivalent to the rest of the class. We did not find any other relationships between scrolling and comprehension.

There was no relationship between scrolling and pre-test scores, reading aptitude or ethnicity. The frequency with which students scrolled roughly corresponded to how much they liked digital environments. Students who preferred reading on digital devices scrolled more than their peers, followed by students who liked both digital and paper reading.

On average, we found that students who scrolled through the shorter section had 27.5 tenth of second timeframes in which they scrolled and students scrolling through the longer passage had 42.7. It seems that about 11 timeframes (1.1 seconds worth) per student are due to the researcher showing them how to scroll at the beginning of the session. If we exclude scrolls in the first minute, the average drops to 16.9 for the shorter section and 31.6 for the longer section. There was a wide range of scrolling behaviors in both conditions. The standard deviation was 19.6 timeframes for the short passage and 29.1 timeframes for the longer passage, with very long tails to the right in both cases.

As you might expect, the amount of time a student spent reading the digital text was positively correlated to how much they scrolled. This correlation was not so strong that we believe scrolling to be an approximation of time, nor did dividing the number of times a person scrolled by the time they spent reading produce different results in our analysis. On average students scrolled once every 14 seconds, but most scrolled more often than that (the median was a scroll every 10.5 seconds).

We also found a correlation between the number of times a student scrolled up and the number of times they scrolled down. This relationship is shown in Figure 4. Initially we thought this might be occurring because too many frames when the screen was still were being classified as scrolls due to bugs in the image processing algorithm. To check this hypothesis, we tried only counting scrolls in which the text moved at least a full line instead of half a line. Using this larger cutoff resulted in some students being labeled as scrolling much less. When we visualized those students, we found that there were many cases where we did not agree that the student had scrolled less than their peers. However, even with the higher cutoff there was a strong relationship between the number of times a student scrolled up and the number of times they scrolled down.

![Figure 4](image)

Figure 4. Number of upward scrolls plotted against number of downward scrolls
**Discussion**
Preferring text to be available in both paper and digital editions is not unusual among adults. A survey of New Zealand college students found that 49.7% preferred that their textbooks be available in both print and online (Traut & Toland, 2014). However, many of the reasons cited by adults for preferring materials in two formats such as reading the digital one on the train and the paper one at home and saving money by reading mostly online and printing critical passages do not apply to middle schoolers. We therefore conjecture that the two fifths of our subjects who did not have a preference between digital and print reading may have not had a chance to develop a preference due to not getting as much exposure to each environment as their peers.

If this conjecture is correct, then it would mean our results line up with Sanchez and Wiley’s (2009) and support the hypothesis that negative effects from scrolling go away as students get more experience scrolling. It is notable that among this population, only down scrolling had a negative relationship with comprehension, not scrolling up. This supports the spatial hypothesis as an explanation of why scrolling is bad for some students. Frequently scrolling down allows the reader to only read the top line of the screen and thus erase spatial cues such as position on the screen which might have helped their memory. Scrolling up indicates that a learner took the opportunity to review or checked the length of the text, neither of which would impact their spatial memory.

One way of adding spatial cues while reading from a screen is to move text by a page at a time so that the spatial cues of the screen are re-introduced. Another method would be to train readers to use other spatial markers like text headings and pictures to orient the text they are currently reading in a space which is not dependent on the current placement of the screen. It is notable that learners who prefer to read from paper scroll less than their peers, suggesting they are employing the first strategy and learners who prefer to read digitally scroll more than their peers, suggesting they are using the latter strategy. Learners without a preference may still be using the screen position to remember what they have read but not adapting their scrolling behavior accordingly. Thus, in this group, those who scroll more are not comprehending the text as well. Training these learners to either scroll less often or pick up on spatial cues like headers and pictures might improve their comprehension.

Historians of printed books have noted that our modern method of reading printed books is only a few centuries old. Sixteenth century English bibles were littered with indices and concordance lists which encouraged discontinuous reading much like the links on Wikipedia sites today (Hillesund, 2010). Annotations and diaries from the era confirm that many readers did take a discontinuous approach to reading their Bibles. In the intervening centuries, long form printed text meant to be read from beginning to end has become popular and methods of reading it efficiently have been developed. As more educational tools move online, it becomes more necessary to train students in practices which will help their digital literacy, which may not be the same as the best practices for reading on paper. This may include introducing them to alternative navigation techniques or different ways of viewing space.

**Future work**
We do not know why there is such a strong relationship between the frequency of downward scrolling and the frequency of upward scrolling. This suggests the metrics for scrolling in this paper are overly broad and a system which could recognize and categorize macro behaviors such as reviewing and checking length would get a better picture of how readers use navigation. Both the statistics we gathered and the visualizations of student scrolling showed in Figures 1 and 2 demonstrate that there is a lot of variance in how middle schoolers navigate scrollable text. In particular, we would like to know whether students who prefer reading digitally employ different navigation strategies than those who prefer paper or do not have a preference. The differences in scrolling amounts we observed between these groups suggest that they are employing different strategies.

**References**


**Acknowledgments**

This work was completed with the support of an interdisciplinary grant from Vanderbilt University (Amanda Goodwin is the primary investigator).
Exploring Predictors of Secondary School Teachers' Use of Technology to Support Student-Centered Teaching

Yaoran Li, Vitality Popov, Veronica Garza, and Anne Keicher
yaoranli@sandiego.edu, vpopov@sandiego.edu, vgarza@sandiego.edu, anne.keicher@gmail.com
University of San Diego

Abstract: The current study aims to explore predictors that independently contribute to high school teacher use of technology in general and for different teaching purposes (student-centered and traditional). High school teachers (N=928) responded to a survey that consisted of measures of their pedagogical beliefs, beliefs towards technology, and perceived training effectiveness. A series of multilevel models were used to explore the independent effects of these factors on teacher use of technology in general and for different teaching purposes. The results showed that teachers’ technology self-efficacy was a significant predictor of teacher use of technology in general. More importantly, teachers’ instructional approach and openness towards technology were more salient when predicting teacher use technology to support student-centered teaching. Our findings suggest that teachers’ pedagogical readiness is as important as technological readiness for technology integration in teaching to serve more advanced teaching purposes.

Introduction
Over the past several decades, technology implementation in schools has been a major reform effort (e.g., Berrett, Murphy, & Sullivan, 2012). Conversations in the United States around transforming teaching and learning via the use of technology is being fueled at the national level by initiatives such as the Call to Action and P21’s Framework for 21st Century Learning (see Office of Educational Technology, 2010 for review). One objective is to use technology to prepare students to be critical thinkers, problem solvers, communicators, and innovators (P21 Partnership for 21st Century Learning, 2009). However, reaching this objective depends on a range of conditions and factors associated with the teacher, student, technology itself, technology-enhanced innovation, policy/legislation, and district/school–level (Groff & Mouza, 2008; Spector, 2010). While all these factors are important for successful technology integration, the teacher who serves as the “innovator” appears to play a crucial role in making pedagogical transformation regarding the use of technology in their teaching (Tondeur et al., 2016). Current evidence indicates that, despite the increased availability of technology in schools (Bulman & Fairlie, 2016), effective integration of technology into teaching and learning, meaning the teacher uses technology as a tool to enhance students’ experiences in the classroom, continues to be a challenge (Inan & Lowther, 2010; Rodriguez, Nussbaum & Dombrovskia, 2011).

Teachers use technology for various purposes to support: 1) administrative or management activities, such as tracking students’ grades, 2) traditional or teacher-centered instructional practices, such as lecturing or presenting, and 3) support student-centered teaching activities, such as giving students choice on how to demonstrate their learning (Palak & Walls, 2009). However, many studies on technology integration or teacher use of technology did not differentiate the purposes of using technology and used the frequency of teacher use of technology tools in the classroom as the outcome measure of technology integration (Aldunate & Nussbaum, 2013; Mumtaz, 2000). Only a few studies examined the predictors of teacher use of technology for student-centered practices (Fu, 2013). This limits our ability to understand some important issues, such as whether the predictors of teacher use of technology for teacher-centered instructional practices are the same as the predictors of teacher use of technology to support student-centered teaching practices. To help fill this void in the existing literature, the present study explored the independent contribution of a variety of teacher factors in predicting their use of technology tools: in general, to support traditional instructional practices, and to support student-centered teaching purposes.

Theoretical framework and significance of the work
Technological Pedagogical Content Knowledge (TPACK, Koehler & Mishra, 2009) highlighted that teachers’ content knowledge and pedagogical competencies are equally important as the technology capabilities, and suggested that a thoughtful alignment of three interconnected capabilities can support teacher effective integrate technology into teaching practice (Voogt et al., 2013). Besides these teacher beliefs and knowledge, International Society for Technology Education (ISTE) revealed that the development of TPACK requires ongoing
organizational support and professional development with dedicated time for teachers to (re)design and enact technology-enhanced lessons (ISTE, 2009). In alignment with these theoretical frameworks, a considerable amount of empirical research studies has been published over the past two decades to explore factors that influence teacher use of technology tools in teaching (e.g., Ertmer, 1999, 2005; Hew & Brush, 2007; also see Buabeng-Andoh, 2012, for review). Although teacher-related factors have been widely viewed from different perspectives, there are mainly four strands of research in the literature, which are: 1) teachers’ attitudes or beliefs towards technology, 2) teachers’ pedagogical beliefs, and 3) teachers’ perceived training effectiveness and organizational support. Below, we reviewed some major findings in the existing literature on these factors.

A considerable amount of research has been conducted to investigate the influence of teachers’ beliefs towards technology on their use of technology (see Buabeng-Andoh, 2012, for a review). Teachers’ perceived competency beliefs of technology, or self-efficacy in using technology, has been found to relate to a more frequent use of technology in the classroom. Teachers with prior computer experience are more likely to learn new necessary skills, such as looking up information more quickly and seamlessly than those who have no prior experience (Groff & Mouza, 2008). On the contrary, teachers lacking confidence in their computer skills are less likely to use technology into their teaching practices (Wozney, Venkatesh, & Abrami, 2006). In addition to perceived competence in technology skills, teachers’ positive attitude toward technology, such as passion about technology, openness towards technology, or feeling comfortable using technology, may also affect their technology integration practices (see Hew & Brush, 2007, for a review).

Research has also suggested that teachers’ pedagogical beliefs are an important predictor of their use of technology (e.g., Ertmer & Ottenbreit-Leftwich, 2010; Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012). For example, Hermans, Tondeur, van Braak, and Valcke (2008) examined the influence of teachers’ educational beliefs on teacher use of computer with 525 primary school teachers. They found that teachers’ constructivist beliefs predicted unique variance in teacher use of computers above and beyond teachers’ background variables and teachers’ attitudes towards computers. As part of a large-scale national study in the Netherlands, Drent and Meelissen (2008) revealed that primary and secondary teachers’ background, technology competency beliefs, attitudes toward technology, and their pedagogical approach all explained unique variance in their use of technology to support educational objectives. Based on findings from a qualitative approach, Ertmer et al., (2012) suggested that teachers with student-centered beliefs tended to use technology through a more student-centered approach. However, research in this area is still limited comparing to research in teachers’ beliefs towards technology.

Some research suggests that teacher trainings are related to teacher technology use. For example, in an empirical study that employed the Teacher Attribute Survey, Vannatta and Fordham (2004) suggested that the amount of technology trainings teachers received is a predictor of teachers’ technology use. Another institutional factor that have been also suggested important is organizational supports, which means to provide teachers with time and environment to practice the ways to integrate technology in teaching and getting feedback. Based on a longitudinal case study, Levin and Wadmay (2008) suggested that opportunities to practice, reflect, and interact with other teachers are crucial in the process of facilitating classroom technology adoption. Also, Wong and Li (2008) found that the collaborating and an experimentation culture set by school leaders influenced effective technology integration. Although it is intuitive to relate trainings and supports focusing on technology with increased technical skills and use of technology, trainings or supports that focus on teachers' curriculum development or pedagogical practices may be as important as efforts to improve teachers’ technology using skills (Sandholtz & Reilly, 2004).

Although most of existing literature focused on predictors of teacher use of technology in general, some recent studies suggested that teachers’ pedagogical beliefs, teachers’ self-efficacy around technology, and received professional development support predicted teachers’ use of technology in student-centered ways (Ananiadou & Claro, 2009; Chen, 2010; Miranda & Russel, 2012). These studies, however, did not differentiate and compare the predictors of teachers’ use of technology for different purposes. Another layer of complexity in this research area relates to inclusive investigation of multiple teacher-related factors and their independent predictive effects. As noted earlier, only few studies investigated the independent contribution of teachers’ pedagogical beliefs, teachers’ technological beliefs and attitudes, and teacher training to teacher use of technology (see Tondeur et al., 2016 for review). Also, most studies focused on pre-service teachers and primary school teachers, thus, research on high school teachers is limited.

Considering these research gaps, the current study aims to include multiple types of teacher-related factors (pedagogical beliefs, technological beliefs, and perceived professional development) that can help us understand the independent contribution of each set of predictors in terms of using technology for different teaching purposes. A large set of survey data were collected from high school teachers from a large, urban school district implementing a one-to-one technology initiative. For this study, the following research questions are...
addressed: 1) how are the teacher-related factors independently predict high school teacher using technology in general? 2) how are the teacher-related factors independently predict teacher use of technology to support student-centered teaching practice? 3) how are the teacher-related factors independently predict teacher use of technology to support traditional teaching practice?

Methodologies

Participants and procedure
As part of a larger longitudinal study conducted in a large, urban K-12 public school district in the Southwestern United States, a district-wide teacher survey was disseminated to high school teachers in spring 2016 with an online survey system. Researchers received teacher emails from the district and all high school teachers were sent an individual email via their school requesting their participation. The survey remained open for two weeks and during this time, teachers were sent three reminders about the survey but the survey participation was voluntary. In this school district, approximately 75% of students are classified as economically disadvantaged, and the majority of students are Hispanic (62%) and African American (25%) in 2016 as indicated by indices on the district’s website. All high school teachers received the survey link, and participation was anonymous and voluntary. Within a two-weeks survey window, 1,054 high school teachers answered on survey from 38 high schools, a 52% return rate. Of the 1,054 teachers who participated, 928 respondents passed at least one of the fraud items. Fraud items are commonly used in online survey to ensure participants answer the questionnaire seriously and carefully. For example, the questionnaire stated the clear directive to “Please choose Strongly Disagree on this item.” If a participant answered anything other than “Strongly Disagree,” then the participant did not pass that item. Survey data of the 928 participants were included in the current study. Of the 928 participants, 26% were English teachers, 19% were math teachers, 20% were science teachers, 18% were social science teachers, and 17% were teaching other subjects. This sample included more female teachers (59%) than male teachers. Forty-two percent of the teachers had more than 10 years of teaching experience, and 28% of the teachers had less than three years of teaching experience. About 39% teachers were 24-34 years old; 26% were 35-44 years old; and the remaining teachers were 45-years-old or older.

Measures
The survey consisted of a variety of measures on teachers’ perceptions and practice related to their technology use and classroom instruction. For the current study, the variables that we were interested in serving as predictors were teachers’ pedagogical beliefs, teachers’ attitudes or beliefs towards technology, teachers’ perceived training effectiveness and organizational support, with teachers’ demographics included as covariates. The outcome variables were teachers’ use of technology for different purposes. Most survey items related to teacher use of technology were adapted from the Second Information Technology in Education Study Teacher Questionnaire (SITES, 2006) and the Technology - Instructional Practices Survey for Minnesota Teachers (Minnesota Department of Education, 2014), and some were developed by the researchers. Questions related to teachers’ pedagogy were developed by the researchers based on multiple theoretical frameworks such as the Bloom's Revised Taxonomy Model (Anderson & Krathwohl, 2001), 21st Century Skills (P21 Partnership for 21st Century Learning, 2009), and ISTE Standards for Educators (2008). The researcher-developed survey items were created and refined through an iterative process between researchers specialized in educational technology and teacher education and consultants with expertise in in quantitative methods and assessment. An early draft of the survey was created and piloted by the researchers in spring 2015, and 519 high school teachers from different content areas participated in the pilot study. The researchers slightly revised the survey and then piloted the survey again with a group of 14 high school teachers before it was administered for the current study. The researchers revised the survey items one more time based on feedback from the teachers to ensure that the survey items about teachers’ beliefs and actual practice were relevant. Researchers also revised the item of the teacher use of technology to have a clear focus on student-centered learning versus traditional teaching methods.

Analyses
For the first research question, we explored whether the variables of interest are predictive of teacher use of technology tools in general after controlling for teachers’ background variables. For the second research question, we explored to what extent teachers’ pedagogical belief, teachers’ technological belief and attitude, and teacher training predicted teacher use of technology to support traditional teaching purpose and teacher use of technology to support student-centered teaching purpose after controlling for teachers’ background variables and teacher use of technology tools in general. With teacher use of technology tools in general being controlled, the analysis can
identify the independent contribution of the targeted predictors to teacher use technology for different teaching purposes. Preliminary data analysis showed that the random effects between schools are significant in the null models when predicting teacher use of technology in general (p <.01) and teacher use of technology to support traditional teaching (p = .02). The intraclass correlation coefficients (ICCs) of the null models predicting teacher use of technology in general, teacher use of technology to support student-centered-teaching, and teacher use of technology to support traditional teaching were .09, .02, and .05, respectively. To control for the random effects between schools and to have a consistent structure of the models, a random intercept was added to all three prediction models using PROC MIXED (SAS, 2004). A series of hierarchical linear models were used to investigate these two research questions. All data available for the modeling were used in this study. For each research question, we ran a model with only the teacher background variables first, and then a model added in all the personal and institutional predictors. Except for the teacher background variables, all the predictors and outcome variables were standardized using PROC STANDARD (SAS, 2004) with M = 0 and SD = 1. In these models, female teachers were used as the reference group when investigating the gender difference, teachers who were 45-years-old or older were used as the reference group when investigating the age difference, teachers who taught science were used as the reference group when investigating the content difference, teachers who taught lower grade levels (9th and 10th grade) were used as the reference group when investigating the grade level difference, and teachers who had more than 10 years of teaching experience were used as the reference group when investigating the teaching experience difference.

**Results**

Table 1 shows the descriptive statistics of all the predictors and outcome variables and the correlations between them. The correlation between the two pedagogical beliefs variables (Learning Goals, Instructional Approaches) was moderate (r = .40); the correlation between the two attitudes towards technology variables (Openness to Technology to Support Instructional Practice, Self-Efficacy in Using Technology) was moderate (r = .27); The correlation between the two training variables was high (r = .59). All the predictors were correlated with teacher use of technology tools in general to a small size (rs = .14 -.24), and the largest correlations appeared with the attitudes towards technology variables (rs = .23, .24). Regarding teacher use of technology to the support student-centered teaching purposes, most of the predictors correlated with this outcome to a medium size (rs = .22 -.38). In contrast, the pedagogical beliefs variables (rs = .15, .18) and the training variables (rs = 19-.20) had lower correlations with teacher use of technology to support traditional teaching purpose than attitudes towards technology variables (rs = .23, .26).

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics and correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Goals</td>
</tr>
<tr>
<td>Approaches</td>
</tr>
<tr>
<td>Openness</td>
</tr>
<tr>
<td>Efficacy</td>
</tr>
<tr>
<td>Training</td>
</tr>
<tr>
<td>Training</td>
</tr>
<tr>
<td>Tech G</td>
</tr>
<tr>
<td>Tech S</td>
</tr>
<tr>
<td>Tech T</td>
</tr>
</tbody>
</table>

Note: Goals: Learning Goals; Approaches: Instructional Approaches; Openness: Openness to Technology to Support Instructional Practice; Comfort: Comfortable with Technology to Support Instructional Practice; Efficiency: Self-Efficacy in Using Technology; Training T: Effectiveness of Technology-Focused Training; Training G: Effectiveness of General Training; Tech G: Use of technology tools in general; Tech S: Use of technology to support student-centered teaching purpose; Tech T: Use of technology to support traditional teaching purpose.

Table 2 shows the estimates of the predictors when the background variables were controlled as covariates. A preliminary regression analysis showed that none of the predictor has violated the multicollinearity rule (VIFs < 3). When predicting teacher use of technology tools in general (Model 1B), both attitudes towards technology variables, self-efficacy in using technology and openness to technology to support instructional practice were significant (β = .15, p < .01), and (β = .08, p = .07), respectively, indicating that teachers who have higher confidence in their technology skills and who are more open to using new technology in teaching use technology tools more frequently in their teaching practices in general.
When predicting teacher use of technology to support teaching purposes (Model 2B & Model 3B), teacher use of technology tools in general was a significant predictor in both models ($\beta$s = .17, .21, .19, $p$s < .01), as well as teachers’ self-efficacy in using technology ($\beta$s = .11, .13, $p$s = .01). Furthermore, Model 2B shows that teachers’ instructional approach also independently predicted teacher use of technology to support student-centered teaching ($\beta$ = .23, $p$ < .01), however, teachers’ instructional approach was not significant when predicting teacher use of technology to support traditional teaching. Interestingly, openness towards technology independently predicted teacher use of technology to support student-centered teaching ($\beta$ = .15, $p$ < .01) but not traditional teaching. These findings indicate that teachers who frequently use a student-centered teaching approach or who are more open to experimenting with technology are more likely to use technology to support student-centered teaching. Given the exploration nature of this study, with Bonferroni corrections, teachers’ self-efficacy in using technology was the only significant predictor of using technology in general and use of technology to support teacher-centered teaching, while teachers’ instructional approach and openness to technology significantly predicted teacher use of technology to support student-centered teaching.

In addition, although the perceived effectiveness of training was not significant when predicting the three different measures of teacher use of technology, the estimates of the training with a technology focus were marginally significant across three models while the estimates of the teaching training in general were higher when predicting use of technology to support student-centered teaching ($\beta$ = .11, $p$ = .09) than when predicting use of technology to support traditional teaching ($\beta$ = .03, $p$ = .64).

**Discussion**

Student-centered instruction is a highly-valued way to approach learning experiences for students as it focuses on the learner. In a student-centered learning environment, teachers may guide their students through the discovery of new knowledge, facilitate discussions, and/or give students the freedom to explore in their learning. Unfortunately, increased availability of technology in schools has not led to overall improvement in classroom teaching practices (Cuban, 2001; Cuban, Kirkpatrick, & Peck, 2001; Windschitl & Sahl, 2002). The technology tools only helped provide the environmental readiness, or addressed the first-order barriers; how to address the second-order barriers, or overcome the challenges related to teacher readiness (Ertmer, 1999, 2005; Kim et al., 2013) became increasingly important for this stage of technology integration in many schools and districts in the U.S. To address this challenge, one area of research indicates that teachers with more constructivist views and practices tend to not only use technology to support higher order thinking skills, but also use technology more frequently and to support more student-centered curricula (Baylor & Ritchie, 2002; Ertmer et al., 2012; Overbay, Patterson, Vasu, & Grable, 2010). Recent research showed that “tool access”, “constructivist pedagogy”, and a combination of “will” and “skill” can explain a significant amount of the variance (60%-90%) in teachers’ level
of technology integration (Christensen & Knezek, 2017; Knezek & Christensen, 2016; Petco, 2012). While previous studies have examined teacher beliefs and professional development factors as related to teachers’ frequency of technology use, only a few have examined the predictors of teacher use technology in a student-centered way. Furthermore, according to the author's knowledge and a comprehensive literature review, no existing study has differentiated the predictors of teacher technology use for traditional teaching and student-centered teaching. Thus, results of the current study have important implications for education practitioners and researchers.

First, the findings of the current study is aligned with results from previous research on technology self-efficacy (see Hew & Brush, 2007, for a review). When background variables were controlled, technology self-efficacy was still a significant predictor of teacher technology use in general and teachers use technology to support either student-centered or traditional teaching purposes. This result suggests that teachers’ confidence in using technology is directly related to their actual use of technology in the classroom. As in the well-known “Little Engine that Could” story, those who “think I can, think I can” may hold an advantage over those without such beliefs. Bandura (1989) suggested that self-efficacy may influence behavioral outcomes through motivational and affective processes. The mechanisms through which teachers’ technology self-efficacy influence their technology use behaviors may involve these dual processes. Motivationally, as suggested by the TAM model, technology self-efficacy might influence both teaching-related decision-making and later engagement in technology-related instructional activities. Teachers low on technology self-efficacy may correspondingly hold a low expectancy for carrying out the optimal teaching outcomes by using technology, so they may avoid using technology when they do not have to. If technology is asked to be required for their teaching, they may flag in their efforts and work passively. Affectively, technology self-efficacy may enhance teachers’ coping in the face of obstacles when designing and delivering technology integrated instruction. Teachers who have stronger beliefs may be able to experience less anxiety, think more openly and persist longer when faced with difficulties than teachers who are beset by self-doubt (Bandura, 1989).

Second, one important finding of the current study is, when predicting teacher technology use to support student-centered teaching, teachers’ instructional approach was an important contributor. The effect of teachers’ instructional approach was found to be independent from teachers’ technology self-efficacy. Notably, when predicting use of technology to support student-centered teaching together, the estimate of teacher’s instructional approach was more than double the size of the estimate of the technology self-efficacy, suggesting that teachers’ pedagogical approaches and technology usage tendency are both important, but teachers’ pedagogical approach is even more crucial to determine teachers’ use of technology for more desired learning outcomes. This finding is aligned with the TPACK Model, which has been widely used to provide a foundation for practitioners and researchers to understand the multiple components of supporting teachers in their practice and the relationship to technology (Koehler & Mishra, 2009). The TPACK framework emphasizes the integrated roles of teachers’ technological knowledge, pedagogical knowledge, and content knowledge and suggests that good practice requires all three components. The current study did not focus on one specific content area, and we did not measure teachers’ content knowledge, but our results highlighted the importance of teachers’ pedagogical readiness and technological readiness to effective teaching with technology.

Third, we found that teachers’ openness towards technology was also an independent predictor of teacher use technology to support student-centered teaching when technology self-efficacy was controlled. Although teachers who were more confident in their technology ability may be more willing to experiment and practice, these two constructs are different. Another factor that could influence teachers’ openness to technology is teachers’ mindset. Dweck, Chiu, & Hong (1995) suggested that people’s mindsets about the malleability of ability frame the way they perceive and interpret experiences and events, which in turn influence their reactions and responses in such situations. People with a fixed mindset (also referred to as entity theory) tend to believe that ability is fixed and unchangeable, while people with a growth mindset (also referred to as incremental theory) tend to believe that through effort and appropriate strategies, learners can improve their ability. People with a growth mindset are more likely to focus on skill improvement and effective strategy use rather than documenting ability and superficial strategy use. In a new 1-to-1 program, most teachers are not skilled at using instructional technology; however, teachers who have a growth mindset may be more likely to learn how to improve their skills and take risk to try new technology and pedagogy, while teachers who have a fixed mindset would feel more comfortable using the traditional way to teach and to maintain their performance. The current study did not explicitly measure teachers’ mindset, but the results on openness towards technology suggest that the teacher mindset might have an active role in determining teachers’ choice of teaching approach and their performance over time. Future research can investigate how teacher mindset influences teachers' perceptions, practice, and performance regarding instructional use.
This study has important implications for practitioners organizing professional learning experiences for teachers. Research in professional learning has highlighted issues of self-confidence and coaching as important and difficult. For example, Goff and Mouza (2008) suggested that teachers lack of computer knowledge and experience is the most foreseeable challenge for teachers implementing instructional technology in the classroom. Our research confirmed that teachers’ confidence in using technology is a starting point for teacher use of technology, either for traditional teaching use or for student-centered teaching purposes. More importantly, our findings suggest that developing systems of both technological and pedagogical support that accommodate teachers’ technology and pedagogical skills may help teachers integrate technology into their classrooms more effectively. Only technological support is insufficient to equip teachers with the skillset to implement technology to create a student-centered learning environment. Furthermore, building a culture that embraces innovation and experimentation with new technology may be also important. Effective professional development needs to address school culture, teachers’ mindset, and provide sufficient time for modeling, experimentation, and reflection, as well as follow-up support for technology integration in the classroom.

References


STEAM Learning in an In-school Makerspace: The Role of Distributed Spatial Sensemaking

Kay E. Ramey, Reed Stevens, and David H. Uttal
kay.ramey@northwestern.edu, reed-stevens@northwestern.edu, duttal@northwestern.edu
Northwestern University

Abstract: This study examines technology-enhanced STEAM learning among fifth and sixth grade students in one set of in-school makerspaces. It focuses on the learning of one set of meta-disciplinary skills, spatial skills. Prior research has shown these skills to be relevant for STEAM achievement, but they have been underemphasized in our schools and in the literature on learning through making. Informed by a distributed cognition perspective and using a combination of qualitative categorical coding and interaction analysis, this study provides a learning sciences approach to studying spatial thinking and learning. Analyses show that during making activities students engaged in frequent and diverse spatial thinking with a variety of social and material resources and that the sociomaterial contexts of different making activities facilitated different types of spatial thinking. They also show that spatial thinking developed over time and led to problem-solving insights.

Keywords: STEAM, makerspace, spatial thinking, interaction analysis, mixed methods

Makerspaces have become explosively popular in recent years. Many believe they hold promise as contexts for integrated STEAM (science, technology, engineering, arts, and math) learning, meta-disciplinary skill learning, and promoting interest and equity in STEAM (e.g., Blikstein, 2013; Hilton, 2010; Martin, 2015; Sheridan et al., 2014; Vossoughi & Bevan, 2014; Vossoughi et al., 2013). However, we still know relatively little about what is actually learned in these spaces, how it is learned, and how to evaluate learning in ways that don’t interfere with the informal structure of making activities. As makerspaces gain in popularity and move increasingly from informal contexts into schools, it is essential that we answer these questions.

This paper tackles these questions, using one particular set of skills, spatial skills, as an example. We’ve chosen to focus on this particular set of skills, because spatial skills predict performance in college STEAM courses (e.g., Hsi, Linn, & Bell, 1997; Sorby, 1999; Sorby, 2009; Sorby & Baartmans, 2000; Sorby, et al., 2013; Tseng & Yang, 2011) and entry into STEAM disciplines (e.g., Humphreys, Lubinski, & Yao, 1993; Lubinski, 2010; Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009). They also play a central role in the practices of STEAM professionals (e.g., Dogan & Nersessian, 2010; Stevens & Hall, 1998), and are often used in everyday thinking and learning (e.g., Hutchins, 1995; Scribner, 1984; Wagner, 1978). Further, contrary to the historically dominant notion that spatial skills are innate and fixed, recent research has demonstrated that these skills are learnable (Uttal et al., 2013) and thus can be improved through instruction or hands-on experience. However, traditional, textbook learning often de-emphasizes spatial thinking, in favor of verbal or analytic approaches to knowledge. As a result, spatial skills are systematically undervalued and underdeveloped in our schools (e.g., NRC, 2006; Newcombe, Uttal, & Sauter, 2013).

In contrast, hands-on, project-based, learning activities, like the ones found in makerspaces, have the potential to spatialize (Newcombe, et al., 2013) STEAM content, by situating learning within work with physical and digital objects and spatial representations, rather than limiting it to the verbal and analytic domains. In particular, the technological tools used in makerspaces, such as CAD (computer aided design) software and 3D printers, both require and have the potential to improve spatial skills (Basham & Kotrlik, 2008; Onyancha, Derov, & Kinsey, 2009; Shavalier, 2004; Sorby, et al., 2013). However, analysis of the development of these skills is conspicuously absent from the literature on learning in makerspaces. This leaves open questions regarding how or whether these skills are actually used or learned in these spaces and how they support other types of learning. This paper addresses these questions using data from an ethnographic study of middle school students engaged in STEAM learning activities in one set of technology-rich, in-school makerspaces.

Theoretical framework

We know relatively little about how spatial skills are learned and how they might support early STEAM learning, either in makerspaces or in other STEAM learning contexts. This is, in part, because the research that has been done on spatial thinking and learning has relied primarily upon psychometric assessments or laboratory experiments and exogenous rather than endogenous (Hall & Stevens, 2015; Stevens, 2010) accounts of learning.
Further, efforts at improving spatial thinking have focused on didactic, teacher (or experimenter) led instruction on specific, narrowly-constrained topics, rather than on ways to support student inquiry and problem-solving (e.g., Atit et al., 2016; Congdon & Levine, 2017; Novack et al., 2014; Stull & Hegarty, 2016).

In contrast, learning in makerspaces is more student- and inquiry-driven and spans a wide variety of tools and concepts. This necessitates a shift in focus away from didactic instructional approaches and standardized assessments and toward: (1) understanding the ways in which students spontaneously engage in spatial thinking and problem-solving, using a variety of social and material resources; and (2) designing activities that provide the right resources and task constraints to encourage particular types of spatial thinking.

To fill this gap, we draw on distributed theories of thinking and learning (e.g., Goodwin, 2000; Hutchins, 1995; Latour, 2005; Ramey & Uttal, 2017; Stevens & Hall, 1998) to frame our investigation of spatial thinking in makerspaces. Based on this theoretical framing, we conclude that spatial thinking should be examined within the sociomaterial context in which it is learned and applied. We also conclude that we should: (1) examine the specific interactions between people, tools, and representations through which spatial thinking is enacted and developed; and (2) trace specific spatial representations as they traverse across representational media, in order to understand how spatial understandings are distributed to or co-constructed by learners and their sociomaterial context. In doing so, we draw on the construct of distributed spatial sensemaking, or the idea that learners both employ cognitive spatial processes and draw on context- and activity-specific social and material resources to co-construct understandings of spatial phenomena (Ramey & Uttal, 2017). Here, we expand upon prior work by exploring: (1) the role that specific technological tools and activities available in makerspaces (e.g., CAD software, 3D printers, electronic circuits) play in facilitating specific types of spatial thinking; (2) how participation in making improves spatial thinking over time; and (3) the role that spatial thinking plays in supporting other forms of STEAM thinking and learning.

**Method**

The research presented here was conducted in one set of in-school makerspaces, FUSE Studios (Stevens et al., 2016). FUSE provides students with a set of integrated STEAM making and design challenges. These challenges are designed to be interest-driven, learner-centered, and inclusive of many different types of learners. There are almost 30 challenge sequences in which students complete challenge levels of increasing difficulty, according to their interests. Guidelines and resources for FUSE challenges are housed on the FUSE website (https://www.fusestudio.net). However, the actual challenges are done using a combination of open-source software programs, such as Sketchup or Inkscape, housed on students’ local computers, and physical tools and materials, such as 3D printers, circuit boards, or building materials, stored in individual FUSE studios.

The research presented here was conducted in five fifth and sixth grade classrooms, from one large, suburban, school district, with a relatively racially and socioeconomically diverse student population. Of the 127 students in the five focal classrooms, 90 agreed to participate in this research. Of these, 58 were fifth graders, and 32 were sixth graders. 42 were male, and 48 were female.

Drawing on cognitive ethnographic methods (Hutchins, 1995; Hollan, Hutchins, & Kirsh, 2000), we conducted ethnographic observations of classroom activity in five FUSE studios, one during the Spring of the 2014-15 academic year, and the other four during the entire 2015-16 academic year. For these observations, I, or another a member of the research team, attended every FUSE session (two per week for a total of 90 minutes), and collected field notes, video recordings, and pictures of artifacts. Video was collected using one tripod-mounted, stationary camera, and six point-of-view cameras (small Go-Pro®, Drift*, or Mobius® cameras mounted on tennis visors), worn by six focal students in each class.

Initial content-logging of the video data showed that of the 24 different FUSE challenges available to students during our observations, we had adequate documentation of students doing 18 of them (at least two students or groups of students doing the challenge over one or more classes). For these 18 challenges, we selected for analysis two, contrasting cases of a student or group of students doing the challenge. We selected the first student case based on the amount and quality of available video, privileging cases where students worked most or all the way through the challenge. In selecting the second student case from each challenge, we chose a case that contrasted with the first case along one or more theoretically important dimensions (e.g., individual versus collaborative, fifth versus sixth grader, or systematic versus tinkering approach). For each case, we analyzed all of the video data of that student doing the given challenge (anywhere between 30 min and fifteen hours of video per case).

In analyzing this data, we used a combination of qualitative categorical coding and interaction analysis. We used a modified version of Ramey and Uttal’s (2017) coding scheme to code multimodal idea units for evidence of distributed spatial sensemaking. This included coding talk, gesture, and object manipulation for evidence of cognitive spatial processes. Codes for cognitive spatial processes were derived from a recent
taxonomy (Newcombe & Shipley, 2015; Uttal, et al., 2013), which compiles the diverse array of spatial skills identified in literature and divides them into intrinsic-static (e.g., disembedding), intrinsic-dynamic (e.g., mental rotation), extrinsic-static (e.g., locating an object or self with respect to a frame of reference), and extrinsic-dynamic (e.g., perspective taking). For the purposes of this study, we also revised this coding scheme, based on our observations of the types of spatial thinking students were engaging in. Drawing on prior work by Hutchins (1995), Goodwin (2000), and Stevens and Hall (1998), we also coded participants’ idea units for both the human and non-human resources they were drawing on to aid in spatial sensemaking and problem-solving, including diagrams, instructional videos, written instructions, other students’ descriptions (multimodal), instructors’ descriptions (multimodal), and tinkering with or exploring materials. Then, we also analyzed episodes of distributed spatial sensemaking, using interaction analysis (e.g., Goodwin, 2000; Hall & Stevens, 2015; Jordan & Henderson, 1995; McDermott, Gospodinoff, & Aron, 1978; Mehan, 1982). We did this to better understand not just what cognitive processes and practices were being used, but how they were being used in interaction and what it was about the specific sociomaterial conditions and task constraints of the FUSE activities that led to their use.

Findings

Our analyses yielded four findings related to spatial thinking and STEAM learning in the context of FUSE activities. First, in making sense of and working through FUSE challenges, students engaged in frequent and diverse forms of spatial thinking and drew on a variety of both social and material resources, often in coordination with one another. Further, the different sociomaterial contexts and task constraints of different FUSE challenges facilitated different types of distributed spatial sensemaking. Over time, the spatial thinking occurring during different FUSE challenges developed, and finally, that spatial thinking led to STEAM problem-solving insights.

Students engaged in frequent and diverse spatial thinking with a variety of resources

Across all the data we analyzed of students working through FUSE challenges, we found 9393 instances of spatial thinking demonstrated through talk, gesture, or object manipulation. Students engaged in 13 different types of spatial thinking, spanning all four quadrants of the two by two taxonomy. The most commonly demonstrated were extrinsic-static skills (57 percent of total instances of spatial thinking), including thinking about spatial relations between objects or between self and objects (54 percent) and describing relative size (3 percent). These were followed by intrinsic-static skills (24 percent), including disembedding (17 percent), quantifying space (5 percent), and categorizing space (2 percent), then extrinsic-dynamic skills (11 percent), including perspective-taking (6 percent) and thinking about dynamic spatial relations between objects (5 percent), and finally intrinsic-dynamic skills (8 percent), including mental rotation (3 percent), 2D to 3D translation (3 percent), scaling or scale changes (1 percent), mental simulation (1 percent), and mental folding (less than 1 percent).

There are two things that are important to highlight in these findings. The first is the very large amount of spatial thinking going on during these activities (9393 instances). The second is the broad range of different spatial skills students demonstrated, and in particular, the relative infrequency of intrinsic-dynamic spatial thinking (8 percent or 713 instances), relative to other types of spatial thinking. This is important, because most of the psychometric tests used in correlational studies test primarily for these intrinsic-dynamic skills. So, by relying only on those, it’s clear that we’re missing a lot of what’s going on in real-world problem-solving contexts.

Another important aspect of spatial thinking in real-world learning contexts that laboratory and correlational studies fail to account for is the heterogeneity of social and material resources that students draw on to make sense of spatial concepts and how the use of those resources shapes spatial thinking. In making sense of the spatial aspects of the various FUSE challenges, students used a variety of both social and material resources, often in coordination with one another. Social resources included both other students in the classroom (44 percent of total resources used) and the adults serving as FUSE facilitators (9 percent). Material resources included diagrams (3 percent), help videos (10 percent), and written instructions from the FUSE website (7 percent), physical or digital materials specific to challenges (28 percent), and sketches created by students (1 percent).

There are two things that are important to notice in these numbers. First is that many of the resources students drew upon, such as help videos, diagrams, sketches, and physical and digital materials had strong, inherently spatial components, whereas others, such as other students, facilitators, and written instructions, were not inherently spatial but were able to convey spatial information through practices such spatial language, gesture, and object manipulation. The second thing worth noticing is how infrequently students drew on the
facilitator as a resource, relative to other resources available in the classroom. This contrasts with the structure of a traditional school classroom, where the teacher is the primary resource from whom knowledge is dispensed. This difference is important, as it emphasizes the need to move away from those traditional, didactic approaches to studying and improving spatial thinking and learning (or any thinking and learning) and toward looking at students’ own spontaneous thinking and problem-solving with a variety of social and material resources.

Different challenges facilitated different types of spatial thinking

As one might expect, given the role that other people and material resources played in spatial thinking, the different sociomaterial contexts provided by different FUSE challenges led to different types of distributed spatial sensemaking. For example, Figure 1 shows the different spatial skills demonstrated during different challenges.

Figure 1. Spatial skills by challenge, as a percentage of total spatial idea units communicated through talk, gesture, or object manipulation during completion of that challenge.

There are two main things that are important to notice here. First is the relatively high frequency of both intrinsic-dynamic and extrinsic-dynamic spatial thinking in FUSE challenges involving CAD software (e.g., 3D You, Keychain Customizer, Print My Ride, Eye Candy, Dream Home, and Dream Home 2). This indicates the importance of particular technology tools in facilitating particular types of spatial thinking. Second is the relatively high frequency of extrinsic-dynamic skills in challenges, such as 3D You and Get in the Game. We argue that this is because these challenges required the coordinated movement of multiple people, physical and digital representations, and objects simultaneously, in order to complete the challenge.

Interaction analysis of episodes of distributed spatial sensemaking from these challenges demonstrates why this is the case. For example, the transcript in Table 1 shows an interaction between three students, Tia, Kyle, and James, as they worked together to do the last level of the 3D You challenge. The goal of this challenge level is to use a Kinect to scan a 3D image of one student’s head (in this case James’ head) into a software program, so that the student can 3D print a bust of himself. At the opening of the interaction, the students were almost done scanning James’ head. All that was left was to scan the top of it. Kyle had been holding the Kinect, while James, seated in a spinning desk chair revolved slowly in a circle. However, at the opening of this episode, Tia, who had been sitting at the computer monitoring the representation of James' head on the screen offered to switch places with Kyle and hold the Kinect. So Kyle took on the role of monitoring the representation on the computer screen and instructing Tia and James movements in order to finish the scan.

In this excerpt, we can see that in order for the activity to proceed successfully, the participants in the interaction (Kyle, James, Tia, the Kinect, the computer, and the desk chair) needed to both think spatially and coordinate spatial representations across different representational media (gesture, talk, body position, and the computer display). Analysis of the human participants’ contributions to this interaction shows that they engaged in a number of extrinsic-static and -dynamic types of spatial thinking, including perspective taking (lines 14-15,
24-25), and both static (line 6) and dynamic (lines 7, 10, 12, 14-19, 24-15) spatial relations. This episode demonstrates how an activity can be designed to encourage or necessitate the use of particular spatial skills. This contrasts with prior approaches to teaching spatial skills, which have involved didactic teacher-led instruction.

Spatial thinking developed over time

Interaction analysis of spatial thinking during FUSE challenges also led to two important findings concerning learning. First, students’ spatial thinking developed over time, through work on the FUSE challenges. For example, while working on the Dream Home challenge, one student, Johanna, tried to add an extra wing to her CAD model home. However, because when she created it, she was looking at her house from above, she accidentally created it on an angle, rather than flat on the ground. After initially getting frustrated, she decided that she liked the diagonal structure and turned it into a pyramid, which she used as a garage for her home. In fact, she liked it so much that, after accidentally closing her file without saving, she recreated it. In doing so, she was forced to employ perspective-taking to figure out how she had created it in the first place.

Then, in her later work on the Dream Home 2: Gut Rehab challenge, she demonstrated how her ability to engage in perspective taking in the CAD environment (Sketchup) had improved with experience. The illustrated transcript presented in Figure 2 shows how, after an initial plea for assistance from her friend, Victoria, she was able to independently and efficiently use the “orbit” and “pan” tools in Sketchup to select

<table>
<thead>
<tr>
<th>Line</th>
<th>Person</th>
<th>Talk</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Tia:</td>
<td>Okay.¹ Gosh,² how did you stay in this position?</td>
<td>¹((reaches across Kyle’s body and takes Kinect from him without moving it)) ²((holds Kinect in place with right arm outstretched))</td>
</tr>
<tr>
<td>7</td>
<td>Kyle:</td>
<td>¹Now, James, you gotta move a little bit.</td>
<td>¹((walks around table and stands in front of computer, looking at computer screen))</td>
</tr>
<tr>
<td>8</td>
<td>Kyle:</td>
<td>James!</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>James:</td>
<td>What?</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Kyle:</td>
<td>¹Okay, move your chair.</td>
<td>¹((looks at the representation of James on the computer screen))</td>
</tr>
<tr>
<td>11</td>
<td>James:</td>
<td></td>
<td>((turns body and chair slowly))</td>
</tr>
<tr>
<td>12</td>
<td>Kyle:</td>
<td>Okay, come over here a little bit¹</td>
<td>¹((waves hand to the right))</td>
</tr>
<tr>
<td>13</td>
<td>Tia:</td>
<td>Me?²</td>
<td>²((begins moving to her right with the Kinect))</td>
</tr>
<tr>
<td>14</td>
<td>Kyle:</td>
<td>No no no no! Now how do we get the top of his head? Because if you try to pick it up², it just says go back to last pose.</td>
<td>²((holds hand out in “stop” gesture))</td>
</tr>
<tr>
<td>15</td>
<td>James:</td>
<td></td>
<td>((leans forward so they can scan the top of his head))</td>
</tr>
<tr>
<td>16</td>
<td>Tia:</td>
<td>There we go. Now circle around holding your breath.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Kyle:</td>
<td>Yeah, circle around.</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>James:</td>
<td>((begins to turn slowly))</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Kumar:</td>
<td>Ok, slowly, and wait. But this has to be¹ a little bit more.²</td>
<td>¹((waves hand to the right)) ²((points to representation of James’ head on the computer screen, then waves hand to the left))</td>
</tr>
<tr>
<td>20</td>
<td>Computer screen:</td>
<td>(Image of James’ head rotates to side of screen so it looks like he’s sitting on the wall.)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Tia:</td>
<td>¹Whoa!</td>
<td>¹((laughs))</td>
</tr>
<tr>
<td>22</td>
<td>Kyle:</td>
<td></td>
<td>(laughs))</td>
</tr>
<tr>
<td>23</td>
<td>James:</td>
<td></td>
<td>((looks at screen and rotates his head slightly to align with the angle of his head on the screen))</td>
</tr>
<tr>
<td>24</td>
<td>Tia:</td>
<td>What happened? Am I like¹</td>
<td>¹((rotates Kinect back and forth but representation on screen stays the same))</td>
</tr>
<tr>
<td>25</td>
<td>James:</td>
<td>¹Ok, let’s watch the video²</td>
<td>²((gets up and points to the back or exit button on the screen))</td>
</tr>
</tbody>
</table>

Table 1: A Distributed Cognitive System Does the Last Level of the 3D You Challenge
appropriate perspectives on her home to complete her desired task – placing a rug flat on the floor.

Figure 2: Johanna engages in perspective taking to place a rug on the floor of her model home.

Spatial thinking led to STEAM problem-solving insights

The other way in which distributed spatial sensemaking led to learning was by advancing problem-solving. For example, as one group of students worked through the Solar Roller challenge sequence, they encountered a number of problems, many of which were solved through spatial insights. For example, when they began Level 2 of the challenge (which requires that students create a 50-inch-long tunnel for their solar car to drive through), they searched for objects in the classroom that could be arranged into the shape of a tunnel. First, they used chairs, then printer paper, then box tops. Meanwhile, noticing that the wheels of their solar car were spinning faster in midair than on the carpet, they hypothesized that the car would go further or faster on a smooth surface. First, they tried pieces of paper on the carpet; then they moved to a smooth countertop. The students also integrated spatial thinking with mathematical thinking, calculating how many chairs, pieces of paper, or box tops they would need to create a 50-inch tunnel, and measuring the distance their solar car travelled on each run. Finally, throughout the challenge, the students engaged in multiple rounds of troubleshooting and iteration with the solar car itself, using help videos and diagrammatic instructions from the FUSE website or a combination of observations and tinkering to reconfigure their car, all of which involved spatial thinking.

Conclusions and implications

This analysis of learning in FUSE makes four main contributions. First, using spatial skills as an example, it provides a missing, close analysis of learning in makerspaces and thus provides a template for the systematic study of other types of thinking and learning during making activities.

Second, it shows that spatial thinking occurs in making activities, and how it both develops over time and advances problem-solving. As a result, the findings presented here improve our understanding of what is learned in makerspaces and how that learning happens. They also challenge traditional, psychological conceptions of spatial ability as individual, innate, and fixed. Finally, they provide both a missing account of the role of spatial thinking in early STEAM learning and an alternative set of methods for improving spatial skills.

Third, the prevalence of a wide variety of spatial skills and the frequent use of other people, tools, and representations shows what we’re missing from only using psychometric assessments to study spatial thinking and learning. This finding emphasizes the importance of augmenting quantitative, exogenous accounts of spatial thinking and learning with qualitative, endogenous ones (Hall & Stevens, 2015; Stevens, 2010). This shift in methods may be particularly beneficial for capturing the spatial thinking and learning of female and low SES students, who underperform on psychometric assessments (e.g., Eliot & Fralley, 1976; Levine, Huttenlocher, Taylor, & Langrock, 1999; Levine, et al., 2005; Linn & Peterson, 1985) but may be capable of more when placed in more authentic problem-solving contexts and allowed access to relevant social and material resources.

Finally, this analysis shows how the different sociomaterial contexts and task constraints of different activities facilitate different types of distributed spatial sensemaking (e.g., intrinsic-dynamic spatial thinking with CAD software or extrinsic-dynamic spatial thinking in 3D You). This provides an understanding of the types of tools, social arrangements, and activities that facilitate different types of spatial skills. As a result, this analysis can serve as a guideline for educators in selecting making activities to facilitate particular types of spatial skills. It also improves our understanding of the advantages of providing students with access to particular technology tools available in makerspaces, such as 3D printers and CAD software.
References


**Acknowledgments**

This work is supported by the National Science Foundation, under NSF grants DRL-1348800 and DRL-1433724. However, any opinions, findings, conclusions, and/or recommendations are those of the investigators and do not reflect the views of NSF.
Secondary Students’ Evaluation of Inappropriate Strategies of Reasoning About Evidence Under a Scientific Explanation

Guanzhong Ma, The University of Hong Kong, Hong Kong, gzma@connect.hku.hk
Carol Chan, The University of Hong Kong, Hong Kong SAR, ckkchan@hku.hk
Jan van Aalst, The University of Hong Kong, Hong Kong, vanaalst@hku.hk
Jing Wang, The University of Hong Kong, Hong Kong, xiaoyuer19921023@126.com

Abstract: This study explored potential factors that could affect the development of students’ strategies for evaluating scientific evidence. Thirty-six students from Grade 7 were asked to evaluate the evidence from an investigation involving force and motion. They were shown three inappropriate responses to reasoning about that evidence. They were then asked to evaluate these responses against an informational text that conveyed the scientific explanation with Galileo’s thought experiment. Participants’ verbal responses to reasoning about the inappropriate responses to the informational text were transcribed and analysed qualitatively. We found three factors that influenced strategy development: the students 1) kept the conflict between the competing claims of the domain unresolved, 2) doubted the validity of the inference from the thought experiment and 3) mistakenly assessed the plausibility of the competing claims. Implications for explaining the development of reasoning about scientific evidence and for instruction are discussed.

Introduction

Scientists’ use of evidence to evaluate knowledge claims is important in science and science learning. Kuhn (2002) suggested evaluating claims using evidence is the essence of scientific reasoning. The use of evidence is a central component the construction of explanations (Iordanou & Constantinou, 2015).

The purpose of evaluating claims using evidence is to judge whether a claim is correct and how certain this judgment is in light of the evidence. This requires two practices: (1) to examine the tenability of the claim in the face of the evidence (Zimmerman, 2007), which involves determining the extent to which the claim holds with respect to the evidence; and (2) to evaluate the evidence in relation to the claim.

Research purpose and questions

The study focused on the latter practice, that is, evaluation of evidence from scientific investigations, aiming to explain the development of this reasoning. Studies of the development of reasoning during evaluations of claim and evidence have found that a student’s content knowledge in a scientific domain plays a role in his or her reasoning about evidence (e.g., Hergovich et al., 2010). Studies have also found that students’ epistemic or meta-conceptual understanding of scientific knowledge and evidence underlies the development of their reasoning in scientific inquiry (e.g., Klaczynski, 2000; Pluta et al., 2011).

However, most existing studies have focused on the practices of using evidence to evaluate knowledge claims, models and explanations (e.g., Iordanou & Constantinou, 2015; Pluta et al., 2011), rather than the practice of evaluating the evidence itself. Furthermore, research has typically assessed the changes in one’s reasoning to capture the associated strategy development (e.g., Schwarz & White, 2005). Little research has investigated the changes in one’s awareness of or views on the effectiveness of individual strategies during strategy development. Therefore, it remains unclear which factors promote or hinder these changes and thus influence the development of reasoning about evidence in science.

To address this gap in the literature, the study aimed to explore factors that could affect the strategy development of the reasoning about evidence. Specifically, it sought to identify potential factors that could hinder changes in students’ views on inappropriate strategies in the presence of a scientific explanation. This aim was achieved by analysing students’ protocols when evaluating inappropriate responses under the scientific explanation.

The research question was: What are students’ reasons for not judging inappropriate strategies as inappropriate when confronted with a scientific explanation for evidence? This research question asked how students responded to the scientific explanation when evaluating responses that used inappropriate strategies. Specifically, this question sought to identify the reasons why students did not change their views on the inappropriate strategies, even though the scientific explanation contradicted those responses.
Significance
We contend that the practice of evaluating evidence should be a central component in science instruction that features evidence-based inquiry. The use of appropriate and sufficient evidence to support claims has been emphasised as a key dimension of the practice of developing evidence-based explanations (Gotwals & Songer, 2013). Instruction that asks students to evaluate their misconceptions of particular domains against evidence without evaluating that evidence may encourage “blind theory change” (Chinn & Brewer, 1993).

However, as discussed, there is little understanding of which factors affect the development of reasoning strategies, especially with respect to students’ awareness or views on more or less effective strategies. This research can contribute to this understanding by identifying factors that may prevent students from considering inappropriate strategies as less adequate. Pedagogically, this study can suggest the knowledge and understanding that should be emphasised in instruction to promote students’ reasoning.

Methods
A sample of secondary students were asked to evaluate the evidence from an investigation of force and motion and to evaluate a set of given inappropriate responses to this evaluation. Participants were then asked to evaluate the responses against an informational text that conveyed the scientific explanation for the evidence from the investigation. The transcripts of students’ verbal responses when evaluating the inappropriate responses against the scientific explanation for that evidence were analysed to identify their reasons for not viewing the inappropriate reasoning strategies as inappropriate.

Research context and participants
The study took place in an urban private secondary school in a medium-sized city in South China. 36 Grade 7 students (20 boys and 16 girls aged 12 to 13 years) participated in the study. As confirmed by their science teachers, these students had not received formal instruction on the relation of force and motion. The students did have some experience of designing and implementing science experiments (e.g., investigating the factors influencing plant growth) prior to the study. But their teachers suggested they had no experience in evaluating evidence from scientific investigations.

These students were selected from a total of 104 students (including 50 girls) because they made a particular misconception prediction about an object’s movement in a conceptual question. We purposefully selected these students to prevent one’s initial domain theories from influencing the reasoning about evidence. This question showed a scenario in which a wooden block was moving on a frictionless horizontal table. The participants were told that a student was initially pushing the block in the same direction of its movement and then released it. They were asked to predict how the block would move after release by choosing one of six given options.

Materials and procedure
These selected participants worked with the researcher individually in a single session consisting of three parts: a) initial reasoning about evidence sufficiency, b) evaluating inappropriate responses to this reasoning and c) re-evaluating the inappropriate responses to this reasoning against the provided informational text that conveyed the scientific explanation.

All materials shown to the participants were in simplified Chinese. Audio recordings of all materials were made to facilitate the participants’ comprehension of the materials in case some of the participants had difficulty reading the texts by themselves. The participants read these materials while listening to the audio recordings. The materials were translated into English by the first author.

Initial reasoning about evidence sufficiency
Each individual participant completed a written reasoning task on reasoning about scientific evidence. The task was adapted from studies investigating people’s responses to anomalous data (Chinn & Brewer, 1998; Mason, 2000). In this task, participants were presented with a scientific prediction of the object’s movement and an investigation to test this prediction. They were asked to evaluate the sufficiency of evidence from this investigation to evaluate the scientific prediction.

Scientific prediction. The participants were told that a hypothetical student answered the same conceptual question as they did and chose Option D. Always move constantly. This option was correct according to Newton’s first law and was referred to as the scientific prediction.

Description of an investigation to test the scientific prediction. The participants read a brief description of the experiment to test the previous prediction. Several key design details of the experiment were described. First, the experiment was conducted in the same condition as the experiment in the conceptual question, that is,
any friction and resistance could be ignored. Second, the speeds of the block were measured in 0.1-second intervals for a total duration of 2 seconds and that two trials were conducted. Third, the experimental equipment and instrument were operating well during the experiment.

Data obtained in the investigation. The participants read the data obtained in the experiment, as shown in Figure 1. The data were presented in a chart that showed the speeds over time. In both trials, the speed values remained roughly constant with a variation of around 1% over time. Taking the experimental errors into account, the data from both trials suggested that the object moved at a constant speed.

![Figure 1. Data Presented in the Reasoning Task.](image)

The participants were asked to rate the extent to which they thought the data were sufficient for evaluating the initial prediction of the movement on a 5-point scale and to explain their ratings.

Evaluating inappropriate responses of reasoning about evidence
The participants were then presented with three responses to reasoning about the sufficiency of that evidence. These responses involved using the particular inappropriate strategies for evaluating evidence, which were referred to as inappropriate responses. The participants were asked to rate the extent to which they agreed with each response and to give explanations for their ratings.

The inappropriate responses to reasoning about evidence were modified from those identified in a previous study of the authors. Three hypothetical students – A, B and C – were created to deliver these responses to participants. Student A’s response was “I think after the student let it go, because the force keeping the wood block moving disappeared, so the block would slow down and finally stop.” Student A mistakenly predicted that the block would slow down after release. Student B’s response was “I think the speed would decrease. Two seconds is too short to see the speed decrease. We surely can see the speed decrease after the two seconds.” Student B predicted that the block would slow down and considered that the duration for which the speed was measured was too short to observe the block’s slowing down. Student C’s response was “I see from the data that the speed of the blocks moves ups and downs, it is sometimes fast, sometimes slow. This is different from Tom’s choice.” Student C mistakenly interpreted the data as suggesting that the speed was always changing and was thus inconsistent with the scientific prediction.

Each response involved making a particular incorrect prediction of the movement. Student A and Student B’s responses made the misconception prediction (i.e., the block would slow down). Student C’s response made the claim of the data misinterpretation (i.e., the speed would be always changing).

Re-evaluating inappropriate responses against the scientific explanation
Each participant was then presented with an informational text. The informational text conveyed the scientific explanation for the scientific prediction with Galileo’s thought experiment. Galileo observed that objects moved farther on smoother surfaces. The inference from this observation was that if the surface were extremely smooth and no friction was present, objects would never stop and always move at a constant speed. This informational text was intended to make the participants consider Student A, B and C’s responses as inappropriate in terms of the claim about the block’s movement.

When re-evaluating these inappropriate responses, each participant was asked to orally explain what they thought of each inappropriate response while referring to the informational text. The participants’ verbal responses to reasoning about the responses to the informational text were audio recorded.
Data coding and analysis
The audio recordings of the participants’ verbal responses were transcribed verbatim. The transcripts were coded to determine which participants considered the inappropriate responses as incorrect, or remained uncertain whether the responses were correct. The participants’ judgments of the inappropriate responses were easily coded because they were explicitly asked to indicate their judgments when they were re-evaluating those responses. Thus it was not necessary to determine the inter-rater reliability of the coding.

We first calculated the frequencies and proportions for the participants’ judgments of each inappropriate response. The transcripts of the participants who did not consider the inappropriate responses as inadequate were selected. These selected transcripts were further analysed qualitatively to identify the categories of reasons that prevented them from recognising the inappropriate response as inadequate.

Results
Table 1 shows the results of analysing participants’ transcripts when they evaluated the inappropriate responses against the scientific prediction and explanation that contradicted the responses. Of all 36 participants, 15 (41.67%) did not consider Student A’s response as incorrect; 11 (30.56%) did not consider Student B’s response as incorrect; 17 (47.22%) did not consider Student C’s response as incorrect. These participants agreed with this response or remained uncertain whether the response was correct.

Table 1: Participants’ Judgments of the Inappropriate Responses against the Scientific Explanation

<table>
<thead>
<tr>
<th>Inappropriate Responses</th>
<th>Judgment of the inappropriate responses</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consider the response as incorrect</td>
<td></td>
</tr>
<tr>
<td>Student A's response</td>
<td>21 (58.33%)</td>
<td></td>
</tr>
<tr>
<td>Student B's response</td>
<td>25 (69.44%)</td>
<td></td>
</tr>
<tr>
<td>Student C's response</td>
<td>19 (52.78%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not consider the response as incorrect</td>
<td></td>
</tr>
<tr>
<td>Student A's response</td>
<td>15 (41.67%)</td>
<td></td>
</tr>
<tr>
<td>Student B's response</td>
<td>11 (30.56%)</td>
<td></td>
</tr>
<tr>
<td>Student C's response</td>
<td>17 (47.22%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>36 (100.00%)</td>
</tr>
</tbody>
</table>

Analysis of the transcripts of these participants’ responses to the informational text found several categories of reasons for not considering the inappropriate response as inadequate.

Keeping the conflict between claims unresolved
Recall that each inappropriate response involved making a particular incorrect prediction of the movement, which contradicted the scientific prediction in the informational text. To judge the adequacy of the inappropriate response, the participants had to recognise the conflict between the competing predictions in the inappropriate responses and the informational text, and to attempt to resolve this conflict.

However, analyses found that participants kept the conflict unresolved. Specifically, three reasons for why they failed to do were identified: the participants 1) made no attempt made to resolve the conflict, 2) attempted to use data to resolve the conflict, but misinterpret the data and not attempt to use the scientific explanation to resolve the conflict; and 3) requested alternative sources to evidence and explanations requested to resolve the conflict. Due to space limitations, this paper only discusses the second theme; the other themes will be discussed at the conference.

Attempt to use data to resolve the conflict, but misinterpret the data and make no attempt to use the scientific explanation to resolve the conflict
The following excerpt demonstrates that the participant recognised the conflict between the claim of data misinterpretation and the scientific prediction and attempted to resolve this conflict by referring to the data. However, the participant ultimately did not resolve the conflict because he misinterpreted the data and did not further rely on the scientific explanation for that evidence to resolve the conflict. The participant was evaluating Student C’s response against the informational text. Recall that Student C mistakenly interpreted the data as suggesting that the speed was always changing.

Researcher: What do you think of Student C’s response according to the informational text?
Participant #12: Based on the text, Student C is definitely wrong. The text says the speed of the block will not change but continue to move forever, but Student C says the speed is sometimes fast, sometimes slow.
Researcher: Sometimes fast, sometimes slow.

Participant #12: I think that Student C is describing the data, but what the scientist says is different from what the data describe.

Researcher: Oh.

Participant #12: So is the scientist wrong, or are the data wrong?

Researcher: You want to make a judgment?

Participant #12: If I look at the data according to the scientist’s view in the text, then the data show ups and downs, ups and downs in the two trials.

Researcher: Oh.

Participant #12: The text says the speed is always constant, which is different from the data. Then they are different.

Researcher: Well, you think they are different.

Participant #12: Yes.

This participant recognised the conflict between the informational text (i.e., the scientific prediction) and the misinterpretation of the data in Student C’s response. The participant referred to the data from the investigation when he recognised the conflict and examined the consistency of the data with the competing predictions of the movement. The participant correctly used the empirical data to judge the conflicting predictions of the movement. It appeared that the participant would resolve the conflict and (mistakenly) judge Student C’s response as correct because he (mistakenly) considered that Student C’s response fit the data. However, the participant could not determine whether the scientist’s prediction or the data from the investigation were correct. Moreover, the participant did not refer to the scientific explanation for the evidence to further consider the conflict, but rather chose to keep the conflict unresolved.

Doubting the validity of the inference from the thought experiment

Recall that the informational text presented the scientific prediction of the movement as the inference from Galileo’s thought experiment. The following excerpt demonstrates that when the participant doubted the validity of the inference from the thought experiment, he did not accept the scientific prediction from the thought experiment and thus could not consider the inappropriate responses, including the misconception prediction, as inadequate. The participant was evaluating Student A’s response against the informational text. Recall that Student A mistakenly predicted that the block would slow down after release.

Researcher: According to the informational text, what do you think of Student A’s response?

Participant #17: Well, according to this text, I think… Well, I still think I am not sure.

Researcher: Can you say it more specifically?

Participant #17: Student A says the force keeping the block moving has already disappeared, so the speed slows down. But the scientist here says that the less the friction is, the farther it moves, indicating that it is the friction that makes the block slow down. And then there is a collision between the two.

Researcher: Oh.

Participant #17: I think the scientist proves what he is saying using four experiments. But he is only inferring from the four experiments. Something inferred could have errors. He is saying that the block is moving at a constant speed forever.

Researcher: And then?

Participant #17: My own thinking is that perhaps the movement gradually slows down after not 1 year, not 2 years, but hundreds of years, thousands of years. I am not sure about what the scientist is saying, so I am not sure.

Researcher: Well, you think he is inferring the movement?

Participant #17: Yes, he is inferring the movement, but he directly says that based on these experiments the block moves at a constant speed, and judges its movement after thousands of years, tens of thousands of years, billions of years. But we do not
have real data to prove that.

This participant recognised the conflict between the prediction in Student A’s response and the scientific prediction in the informational text. He discounted the validity of the inference from the thought experiment and considered the inference from the thought experiment did not lead to a certain judgment about the movement after a long time. The participant suggested physical experiments data (“real data” in his terms) were necessary to conclude the block’s movement for a long duration.

**Mistakenly assessing the plausibility of claims about the movement**

Evaluating the inappropriate responses of reasoning about the evidence against the informational text involved evaluating claims about the movement, including the scientific prediction, misconception prediction and data misinterpretation. To judge the adequacy of the responses, the participants had to judge the scientific explanation as more plausible than the claim of the data misinterpretation and the misconception prediction.

However, the analysis found that the participants mistakenly assessed the plausibility of claims about the movement. Three reasons for why participants did not consider the inappropriate responses as inadequate were identified: the students 1) considered the claim of the misinterpretation of the data as more plausible than the scientific prediction, 2) considered the scientific prediction as implausible and 3) rendered the claim of data misinterpretation plausible. We discussed the third theme here.

**Rendering the claim of data misinterpretation as plausible**

The following excerpt demonstrates that when the students encountered anomalous data that included unexpected variability, they invented explanations to account for the variability to render the pattern of the data as plausible. As a result, the students agreed with the claim of the data misinterpretation and thus hindered the development of the strategy for reasoning about the evidence. The participant was evaluating Student C’s response against the informational text. Recall that Student C mistakenly interpreted the data as suggesting that the speed was always changing and was thus inconsistent with the scientific prediction.

Researcher: What do you think, according to the informational text, of Student C’s response?

Participant #18: Student C’s response is a bit wrong.

Researcher: Oh.

Participant #18: If the block were on an extremely smooth table surface, then its speed would be very even. A rough table surface could cause the friction to be uneven, and would lead to some difference in the speed.

Researcher: What do you mean by “uneven”?

Participant #18: There are some rough places, there are some bumps that could cause the difference in the speed.

Researcher: Oh. It is said that we considered there was no friction.

Participant #18: No friction, but the inertia would result in some small changes.

Researcher: What do you mean by “small changes”?

Participant #18: Although the block is released, its speed will have some vibration, and although the trend is decreasing, there may be some changes in the beginning, and gradually the vibration could become larger and larger.

Researcher: Oh, you think…

Participant #18: The inertia is decreasing, but the decrease could be uneven.

Researcher: Oh.

Participant #18: The decrease could be slow at first and then fast, so perhaps Student C is correct.

This excerpt demonstrates that this participant considered the data misinterpretation as plausible by proposing an explanation to account for the always-changing speed as suggested by the misinterpretation. This participant actually seemed to appeal to the data to make the claim about the movement of the block. When the participant saw the data misinterpretation in Student C’s response, he attempted to explain it by noting that the speed would change from time to time.
Discussion

Factors influencing reasoning about evidence

The above findings revealed three factors influencing strategy development: 1) attempt and competence to resolve conflicts between competing claims, 2) epistemic understanding of thought experiments as a source of scientific knowledge and 3) assessment of relative plausibility of the scientific explanation to misconceptions.

Attempt and competence to resolve conflicts between competing claims

We found that some participants recognised the conflict between the predictions of the movement as indicated in the inappropriate response and the scientific prediction, yet kept this conflict unresolved. This indicates that the attempt to resolve this conflict is necessary for students to judge the inappropriate responses as inadequate and thus essential for developing the strategy for this reasoning. Even when students attempt to resolve this conflict using the empirical data, they may fail to do so because they may misinterpret the data. The competence involved in interpreting data and the attempt to use available explanations as a source for evaluating competing claims are essential for resolving conflicts and are thus important for the developing a strategy for reasoning about evidence.

In addition, students may fail to resolve that conflict when they assume the existence of alternative knowledge sources to empirical data and explanations. Recognising the role of empirical evidence and the explanations for that evidence and claims is important for resolving conflicts between competing claims of the domain, and thus is important for the development of a strategy for reasoning about evidence.

Epistemic understanding of thought experiments as a source of scientific knowledge

We found that some participants doubted the validity of the inference from the thought experiment. They did not accept the scientific prediction from the thought experiment and thus could not consider the inappropriate responses as inadequate. This indicates that when asked to evaluate the inappropriate responses to reasoning about evidence, students’ beliefs about thought experiments as a source of scientific knowledge may affect their judgment of the responses in the face of the scientific prediction based on a thought experiment, and thus may influence the development of a strategy for this reasoning.

This finding adds to the literature on students’ epistemic status of sources of justification for knowledge claims in science. Sandoval and Çam (2011) investigated primary students’ epistemic status of sources of justification for knowledge claims and found that students appealed to empirical data as their first priority to justify a causal claim over plausible explanations. Little is understood about secondary students’ epistemic status in terms of inferences from thought experiments and empirical data from physical experiments. The findings of this study suggest that secondary students appeal to empirical data as their first priority to evaluate claims of movement over thought experiments.

Assessment of the relative plausibility of the scientific explanation to misconceptions

We found some participants did not consider the inappropriate responses as inadequate due to the plausibility of the competing claims about the movement, including the scientific prediction, the misconception prediction and the claim of the data misinterpretation. This indicates the plausibility of the scientific prediction to students may affect their judgment of inappropriate responses to reasoning about evidence and thus influence the development of this reasoning strategy. This finding also indicates that when students encounter anomalous data that include unexpected variability, they may invent explanations that account for this variability to render the pattern of the data as plausible. As a result, students agree with the data misinterpretation, which hinders the development of a strategy for reasoning about evidence.

The research on conceptual change has long contended that the new scientifically accurate understanding needs to be judged as more plausible than the existing misconceptions for conceptual change to occur (Lombardi, Sinatra, & Nussbaum, 2013). The findings of this study suggest that the plausibility of the scientific prediction influences the ability of scientific prediction and explanations to foster changes in students’ misconceptions.

Implications for instruction

The findings suggest to promote reasoning about scientific evidence, students need an enhanced epistemic understanding of scientific knowledge and a conceptual understanding of particular scientific domains. First, science teachers should make the scientific explanation available and plausible to students to promote their reasoning about evidence. As students evaluate the evidence, teachers could ask them to compare the scientific prediction and explanation and their misconceptions of the domain. When students recognise the conflicts
between competing claims about the particular domain, teachers should prompt them to resolve the conflicts by referring to the evidence and to assess the plausibility of the competing claims.

Second, science teachers explicitly teach the epistemic understanding of scientific knowledge. Teachers must help students reflect on the roles of thought experiments in developing scientific knowledge and the relationship between thought experiments and empirical evidence. This could be useful for promoting students’ acceptance of scientific claims that are drawn from thought experiments.

Future research
Future research could investigate whether and how these factors influence the development of reasoning about evidence. For example, in terms of the epistemic status of sources of justification (Sandoval & Çam, 2011), future research could investigate how students view thought experiments as a justification for knowledge claims and the priority they place on this justification compared with other justifications such as empirical data and scientific explanations. Future research could also design an experiment to test whether an enhanced epistemic understanding of the role of thought experiments in developing scientific knowledge brings about positive changes in students’ reasoning about evidence.

Conclusion
This study explored potential factors that accounted for the development of reasoning during the evaluation of evidence. Specifically, it sought to identify the factors that could hinder changes in students’ views on inappropriate strategies in the presence of a scientific explanation.

The results suggest that the attempt and competence to resolve conflicts between competing claims, the plausibility of the scientific prediction and explanation and the epistemic understanding of thought experiments as a source of scientific knowledge influenced the role of the scientific explanation in changing students’ reasoning about evidence. The findings of this study thus contribute to the literature related to the development of reasoning about scientific evidence. We recommend that in the teaching of reasoning about evidence should make scientific explanation available and plausible to students, and address the role of thought experiments in developing scientific knowledge.

References
Flow in Computer-Supported Collaborative Problem-Solving

Gaëlle Molinari, TECFA-University of Geneva & UniDistance, gaelle.molinari@unidistance.ch
Sunny Avry, TECFA-University of Geneva & UniDistance, Sunny.Avry@unige.ch

Abstract: This study investigated the individual experience of flow and its relations with emotions and the perceived socio-cognitive processes involved in a computer-supported collaborative problem-solving task. Participants were asked to play in dyads the video game Portal 2® and then to individually complete three subjective questionnaires (emotion, collaboration and flow). Results showed that flow is related to both the modelling of the partner’s emotions and the perceived mutual engagement in collaborative processes such as information pooling and transactivity. They support the idea that the flow experience should be considered as a means to improve the quality of computer-supported collaborative learning.

Introduction

One of the main current challenges for research in the learning field is to make a shift from a strictly cognitive approach to an approach in which the cognitive and affective (social-emotional) dimensions of learning need to be investigated as intrinsically interdependent. In such an approach, the experiential nature of learning, for instance, emotions that emerged as the result of subjective cognitive appraisal of control and value of the learning situation (Pekrun, 2006), should be examined in relation to the knowledge (co-)construction process. Emotions are seen here as forming “a critical piece of how, what, when and why people think, remember and learn” (Immordino-Yang, 2015, p.8).

In the Computer-Supported Collaborative Learning (CSCL) domain, it is now well accepted to define collaboration as the dynamic interaction between two spaces, a cognitive space dedicated to epistemic activities (co-producing knowledge, resolving the task/problem) and a relational space dedicated to socio-affective activities (maintaining a positive, stimulating and engaging relationship with the learning partners) (Andriessen, Baker, & van der Puil, 2011; Barron, 2003). Andriessen and colleagues argue that understanding what emotions are involved and how they circulate within the group is necessary in understanding how learners work and learn together. In the EATMINT (Emotion Awareness Tools for Mediated Interaction) project (Molinari et al., 2013; 2017), it has been shown that encouraging partners to explicitly share their emotions throughout collaboration (using an emotion awareness tool) positively impacts their perceived effort of modelling each other (mutual modelling; Dillenbourg, Lemaignan, Sangin, Nova, & Molinari, 2016) as well as of building on each other’s contributions (transactivity; Berkowitz & Gibbs, 1983; Molinari, Sangin, Nüssli, & Dillenbourg, 2008). Results also suggest that perceived transactivity and group performance positively correlate with positive emotions like enjoyment and negatively correlate with negative emotions like boredom (Avry, Chanel, Bétrancourt, Pun, & Molinari, 2017). Despite the consensus regarding the importance of considering the collaborative learning experience as a continuous cycle of tensions (and relaxations) at both the cognitive and social-emotional planes (Andriessen et al., 2011), there is still little research that go deep into the dynamics and mutual influence between emotional states, cognitive states and collaborative processes (Mullins, Deiglmayr, & Spada, 2013).

The present research is exploratory and focuses on the individual experience of flow during a computer-supported collaborative problem-solving task. In particular, our main research question concerns the relations between flow, emotions and the perceived socio-cognitive processes involved in collaboration. The rationale for this research is that there is a need to gain a better understanding of the subjective experience of individuals engaged in a collaborative task on the one hand, and on the other, of how this relates to how they perceive the way they work together to solve the problem. Such an understanding is a necessary step toward providing design principles for technology and environments that aim at promoting positive, engaging and meaningful collaborative learning experiences (Riva, Banos, Botella, Wiederhold, & Gaggioli, 2012).

The concept of flow as a framework (Csikszentmihalyi, 1990) is usually used to investigate students’ engagement in the learning process. Flow is defined as an optimal experience in which the person feels simultaneously cognitively efficient, attentionally engaged, motivated, highly interested and happy (Moneta & Csikszentmihalyi, 1996). More specifically, it is characterized by a combination of cooccurrent states, that is, sense of control, cognitive absorption, distorted perception of time, loss of self-awareness (deindividuation), and autotelic experience (heightened enjoyment) (Heutte, Fenouillet, Kaplan, Martin-Krumm, & Bachelet, 2016). Control, absorption, time transformation and loss of self-awareness are described as the cognitive dimensions of flow whereas autotelic experience refers to its affective dimension. Moreover, being in control and totally concentrated are considered both as conditions for experiencing flow whereas alteration of sense of time, loss of reflective self,
and well-being are considered as effects (Heutte et al., 2016). Flow is expected to occur with meaningful learning tasks, that is, tasks that provide individuals with the opportunity to master new challenges and surpass themselves, and also in situations where they perceive themselves as having the necessary skills to perform the task. In complex learning settings, D’Mello and Graesser (2012) showed that a flow state is experienced when a cognitive equilibrium is reached and/or when learners focus on mastery of the learning task. They also highlighted a positive correlation between flow and learning gains. This relation is explained by the fact that flow is experienced as a reward and provides learners with a motivation to persist in the task despite difficulties and obstacles to goals (that is, cognitive desequilibrium moments). Flow can also be promoted in situations where learners experience positive task-related emotions such as enjoyment and curiosity whereas negative emotions such as boredom and hopelessness relate negatively to flow (Boekaerts & Pekrun, 2016).

Research on the experience of flow in computer-supported collaborative/problem-solving is still relatively rare. Flow is investigated as either an individual or a mutual experience (social flow) in group settings (Walker, 2010). Walker showed that the experience of flow is more intense in group activities than in individual activities. This can be explained by the fact that group activities are more challenging/risky and require considerable skills at both task/cognitive and social/relational levels. Moreover, the intensity of joy associated with the flow state in group settings increases with the level of interdependence between group members and is related to the extent to which emotions are shared (e.g. through contagion) during interaction. Results also showed a positive correlation between flow and group performance (Admiraal, Huizenga, Akkerman, & Dam, 2011). This relation is mediated by the degree to which group members share relevant task-related information and also perceive each other as contributing equally to the achievement of the common goal (Aubé, Brunelle, & Rousseau, 2014). All these results support the idea that there is a need for team managers and also for teachers to consider the flow experience as a means to improve the quality of group processes and outcomes.

In the context of this theoretical background, we propose to explore three main questions:

- To which extent is individual flow associated with the degree of interdependence between group members and is related to the extent to which emotions are shared during interaction? (Question 1)
- To what extent is individual flow related to participants’ ability to accurately perceive their partner’s emotions during collaboration? (Question 2)
- To what extent is individual flow related to participants’ perception of the interaction with their partner, i.e. to perceived socio-cognitive involved in collaboration such as grounding, information pooling, trans-activity, consensus building and coordination? (Question 3)

Method

Sixty-four participants (12 women and 52 men; $M_{age}=22.02$ years, $SD=3.49$), grouped in 32 same-gender dyads, participated to this experiment. They were asked to play a collaborative problem-solving video game called Portal 2®. This game was chosen for two reasons. First, it requires both a high individual involvement and the mutual engagement of participants to solve problems in a coordinated way (which corresponds to the criteria of a highly collaborative task; Roschelle & Teasley, 1995). Second, it involves a set of skills necessary to academic success such as problem-solving skills, spatial cognition skills and persistence (see Shute et al., 2014 for a complete description of the cognitive and motivational skills involved in Portal 2®). In this game, players have to find the way out of several closed rooms. More precisely, they have to manipulate objects in their shared environment to open up passages and move forward to a next room until the exit. It is necessary for players to jointly consider what they must do to progress in the game. When a potential solution is found, they have to take the current position of their partner into account and to engage in mutual coordinated actions.

Dyads were randomly assigned to one of 4 conditions. Participants received a combination of two types of biased feedbacks at six times during the game, namely, their dyad’s level of mastery and their dyad’s ranking (among fictive dyads). The research goal in using such feedbacks was to influence control and value appraisals which are assumed to shape task-related emotions (Pekrun, 2006). The purpose of the present study is not to focus on the impact of the control and value feedbacks on flow experience. There were, however, no main effects nor interaction on any of the five flow dimensions ($p > .10$). Regarding effects of control and value feedbacks on emotions and the perceived quality of collaboration, please see Avry et al., 2017).

The members of each dyad were separated in 2 rooms. Both peers were seated in front of a computer equipped with webcam and a BioSemi® electrophysiological system. They were not able to see each other while playing but could communicate orally thanks to microphone headsets. At the very beginning of the experiment, an individual training phase was proposed to familiarize participants with the game. They then performed the task (30 mn). Immediately after collaboration, they were asked to complete 1) an emotion questionnaire, 2) a collaboration questionnaire and 3) a flow questionnaire. These 3 questionnaires are described below.
Emotion questionnaire
It was derived from the Achievement Emotions Questionnaire (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011), and aimed at measuring the emotions experienced during collaboration. Participants rated the intensity of both their own- and their partner’s emotions using a list of 4 activating negative emotions (Anxiety, Anger, Frustration, Shame), 4 deactivating negative emotions (Deception, Hopelessness, Boredom, Sadness), 5 activating positive emotions (Hope, Pride, Joy, Enjoyment, Gratitude) and 3 deactivating positive emotions (Relaxation, Relief, Contentment). Each of these 16 emotions was thus accompanied by two 7-point intensity scales (1=very low or not at all to 7=very high), one for the participants’ emotions and one for their partner’s emotions.

In a dyad with A and B as partners, the accuracy of the mental model A built of B’s emotions, i.e. Model (A, B, emotions) was computed as the absolute difference between Model (A, B, emotions) and Model (B, B, emotions); e.g. difference between the “intensity of Anxiety felt by B estimated by A” and the “intensity of Anxiety felt by B estimated by him/herself”. This difference was calculated for each of the 16 emotions. Four modelling accuracy scores were obtained, two for positive emotions (one for positive activating emotions and one for positive deactivating emotions) and two for negative emotions (one for negative activating emotions and one for negative deactivating emotions). The smaller the score, the more accurate the model.

Collaboration questionnaire
It was inspired from the rating scheme proposed by Meier, Spada and Rummel (2007) to assess the quality of computer-supported collaborative processes. The present questionnaire was designed to measure the participants’ perceptions of both their own engagement and their partner’s engagement in the collaboration process, in particular in five socio-cognitive processes that play a crucial role in the success of collaboration: Grounding (3 items; e.g. “making sure the other has understood you”); Information Pooling (5 items; e.g. “gathering as much relevant information as possible to solve the problem”; “eliciting personal knowledge from the other that could be useful in solving the problem”); Transactivity (2 items; e.g. “building upon the other’s contributions by integrating them into one’s own perspective”); Consensus Building (5 items; e.g. “looking for the best arguments for or against a potential solution”; “trying to convince the other by justifying one’s own proposals for solutions”); and Coordination (5 items; e.g. “structuring by clearly defining subtasks to perform and dividing them among you equally”; “monitoring the time remaining for the task and taking care not to waste time unnecessarily”). For each of these collaborative processes, items were accompanied with 2 types of 7-point frequency scales (1=never to 7=very often), one for the participants’ level of engagement and one for their partner’s engagement level.

Flow questionnaire
We used here the first version of the EduFlow scale (Heutte et al., 2016) designed to measure flow in learning environments. It is a 16-item scale organized into five dimensions of flow, namely: (1) Sense of Control (3 items; e.g. “I felt I was able to cope with the high demands of the task”); (2) Cognitive absorption (4 items; e.g. “I was totally absorbed in what I was doing”, “I felt I was completely focused on what I was doing”); (3) Alteration of the Sense of Time (3 items; e.g. “I did not notice time passing”); (4) Loss of Self-Consciousness (3 items; e.g. “I was not worrying about what my partner thought to me”); and (5) Autotelic Experience (3 items; e.g. “The activity itself gave me a sense of well-being”). For each of these 16 items, a 7-point answer scale (1=strongly disagree to 7=strongly agree) was used.

Results
All statistical analyses were performed using SPSS V.24 (IBM Corporation).

Individual experience of flow during collaboration
Table 1 shows that the five components of flow, that is, Sense of Control, Cognitive Absorption, Alteration of the Sense of Time, Loss of Self-Consciousness and Autotelic Experience, are experienced at moderate to high levels.

Table 1: Means and Standard Deviations for the 5 Dimensions of Flow

<table>
<thead>
<tr>
<th></th>
<th>Means</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of Control</td>
<td>3.99</td>
<td>1.36</td>
</tr>
<tr>
<td>Cognitive Absorption</td>
<td>6.02</td>
<td>1.06</td>
</tr>
<tr>
<td>Alteration of the Sense of Time</td>
<td>5.87</td>
<td>1.28</td>
</tr>
<tr>
<td>Loss of Self-Consciousness</td>
<td>5.29</td>
<td>1.47</td>
</tr>
<tr>
<td>Autotelic Experience</td>
<td>5.07</td>
<td>1.47</td>
</tr>
</tbody>
</table>
We computed correlations between the five flow dimensions. Figure 1 shows a positive correlation between the two flow conditions, i.e. Sense of Control and Cognitive Absorption. Autotelic Experience is positively associated with the two other effects of flow, i.e. Alteration of the Sense of Time and Loss of Self-Consciousness, and also with the two flow conditions. The correlation between Alteration of the Sense of Time and Loss of Self-Consciousness is very weak and not significant ($r = .05$, n.s.). Finally, Alteration of the Sense of Time is positively related to Cognitive Absorption while it does not significantly correlate with Sense of Control ($r = .08$, n.s.). Finally, Loss of Self-Consciousness is the only effect of flow that is not significantly related to any of the two conditions of flow (Sense of Control: $r = .09$, n.s.; Cognitive Absorption: $r = .04$, n.s.).

Figure 1. Relations between the 5 Dimensions of Flow (Absorption=Cognitive Absorption; Sense of Time=Alteration of the Sense of Time; Self-Consciousness=Loss of Self-Consciousness; Autotelism=Autotelic Experience).

Flow and emotions experienced during collaboration

Flow and emotions
Correlations were computed between the five dimensions of flow and the four types of emotions, namely positive activating/deactivating emotions and negative activating/deactivating emotions (see Table 2). Sense of Control and Cognitive Absorption are positively associated with positive activating and deactivating emotions. The correlation between Cognitive Absorption and positive deactivating emotions is, however, marginally significant (weak correlation). Of these two conditions of flow, only Sense of Control is negatively related to negative activating and deactivating emotions. The correlations between Cognitive Absorption and negative activating and deactivating emotions are very weak and not significant.

Autotelic Experience is positively correlated with positive activating and deactivating emotions, and negatively related to negative activating and deactivating emotions. The correlation between Autotelic Experience and negative activating emotions is, however, marginally significant (weak correlation). Loss of Self-Consciousness is positively but marginally related to positive activating emotions (weak correlation). The correlations between Loss of Self-Consciousness and the other types of emotions (positive deactivating emotions, negative activating and deactivating emotions) are very weak and not significant. Of the three effects of flow, only Alteration of the Sense of Time is not significantly related to emotions experienced during collaboration.

Table 2: Relations between Flow Dimensions and Emotions Experienced During Collaboration.

<table>
<thead>
<tr>
<th></th>
<th>Positive Emotions</th>
<th>Negative Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activating</td>
<td>Deactivating</td>
</tr>
<tr>
<td>Sense of Control</td>
<td>$r = .55, p = .000$</td>
<td>$r = .61, p = .000$</td>
</tr>
<tr>
<td>Absorption</td>
<td>$r = .38, p = .002$</td>
<td>$r = .23, p = .07$</td>
</tr>
<tr>
<td>Sense of Time</td>
<td>$r = .15, n.s.$</td>
<td>$r = .13, n.s.$</td>
</tr>
<tr>
<td>Self-Consciousness</td>
<td>$r = .23, p = .07$</td>
<td>$r = .18, n.s.$</td>
</tr>
<tr>
<td>Autotelism</td>
<td>$r = .61, p = .000$</td>
<td>$r = .52, p = .000$</td>
</tr>
</tbody>
</table>
Below, more specific results are listed:

- **Sense of Control** is positively correlated with seven of the eight positive emotions, the strongest correlations being with *contentment* ($r=.61$, $p=.000$) and *pride* ($r=.54$, $p=.000$). The positive correlation with *hope* is not significant ($r=.15$, n.s.). This flow dimension is negatively correlated with seven of the eight negative emotions, the strongest correlations being with *hopelessness* ($r=-.45$, $p=.000$) and *frustration* ($r=-.38$, $p=.002$). The negative correlation with *sadness* is not significant ($r=-.22$, $p=.08$).

- An increased *Absorption* on the task correlates with an increase in experience of *hope* ($r=.46$, $p=.000$).

- **Autotelic Experience** is positively correlated with seven of the eight positive emotions, the strongest correlations being with *enjoyment* ($r=.68$, $p=.000$) and *joy* ($r=.59$, $p=.000$). The positive correlation with *relaxation* is not significant ($r=.19$, n.s.). This flow dimension is negatively correlated with four of the eight negative emotions, the strongest correlations being with *boredom* ($r=-.45$, $p=.000$) and *sadness* ($r=-.38$, $p=.002$). There are no significant correlations with *anxiety* ($r=.10$, n.s.), *deception* ($r=-.21$, n.s.), *frustration* ($r=-.22$, $p=.08$) and *shame* ($r=-.21$, $p=.10$).

### Flow and modelling of partner's emotions

Correlations were computed between the flow dimensions and the four modelling accuracy scores (i.e. one score per type of emotions). Please note that the negative correlations are a sign for high accuracy in assessing emotions in the collaborative partner. As depicted in Table 3, *Alteration of the Sense of Time* is the only dimension of flow significantly related to the accuracy with which participants assess their partner’s positive emotions. The strongest correlation is with the modelling accuracy score for positive activating emotions.

An increase in *Sense of Control* as well as in experience of well-being (Autotelic Experience) is associated with an increase in the accuracy in assessing negative emotions in the partner. The strongest correlations are with the modelling accuracy scores for negative deactivating emotions.

Finally, *Cognitive Absorption* is linked to the ability to correctly assess negative deactivating emotions experienced by the other, *Loss of Self-Consciousness* to the mutual modelling of negative activating emotions.

#### Table 3: Relations between Flow Dimensions and Emotion Modelling.

<table>
<thead>
<tr>
<th></th>
<th>Modelling of Positive Emotions</th>
<th>Modelling of Negative Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activating</td>
<td>Deactivating</td>
</tr>
<tr>
<td>Sense of Control</td>
<td>$r=.20$, n.s.</td>
<td>$r=.10$, n.s.</td>
</tr>
<tr>
<td>Absorption</td>
<td>$r=.11$, n.s.</td>
<td>$r=.08$, n.s.</td>
</tr>
<tr>
<td>Sense of Time</td>
<td>$r=-.32$, $p=.01$</td>
<td>$r=-.28$, $p=.03$</td>
</tr>
<tr>
<td>Self-Consciousness</td>
<td>$r=-.10$, n.s.</td>
<td>$r=.08$, n.s.</td>
</tr>
<tr>
<td>Autotelism</td>
<td>$r=-.16$, n.s.</td>
<td>$r=.006$, n.s.</td>
</tr>
</tbody>
</table>

### Flow and perceived socio-cognitive processes

All flow dimensions, except *Loss of Self-Consciousness*, are significantly positively related to at least two perceived socio-cognitive processes (Table 4 a/b). Flow is more strongly related to the participants’ perception of their partner’s engagement in the collaboration process than to their own perceived engagement.

*Information Pooling* and *Transactivity* are related to both conditions of flow, i.e. *Sense of Control* and *Cognitive Absorption*, and also to two out of the three effects of flow, i.e. *Alteration of the Sense of Time* and *Autotelic Experience*. *Coordination* and *Grounding* are only related to flow conditions (both *Sense of Control* and *Cognitive Absorption* for *Coordination*; *Cognitive Absorption* only for *Grounding*). *Consensus Building* is related mainly to flow effects, more specifically to *Autotelic Experience*. 
Table 4 a/b: Relations between Flow Dimensions and Perceived Socio-Cognitive Processes.

<table>
<thead>
<tr>
<th></th>
<th>Grounding</th>
<th>Information Pooling</th>
<th>Transactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self</td>
<td>Partner</td>
<td>Self</td>
</tr>
<tr>
<td>Sense of Control</td>
<td>$r=.19$,</td>
<td>$r=.20$,</td>
<td>$r=.28$,</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
<td>$p=.03$</td>
</tr>
<tr>
<td>Absorption</td>
<td>$r=.23$,</td>
<td>$r=.26$,</td>
<td>$r=.25$,</td>
</tr>
<tr>
<td></td>
<td>$p=.07$</td>
<td>$p=.04$</td>
<td>$p=.05$</td>
</tr>
<tr>
<td>Sense of Time</td>
<td>$r=.05$,</td>
<td>$r=.03$,</td>
<td>$r=.27$,</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
<td>$p=.03$</td>
</tr>
<tr>
<td>Self-Consciousness</td>
<td>$r=.06$,</td>
<td>$r=.16$,</td>
<td>$r=.001$,</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Autotelism</td>
<td>$r=.14$,</td>
<td>$r=.18$,</td>
<td>$r=.35$,</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
<td>$p=.005$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Consensus Building</th>
<th>Coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self</td>
<td>Partner</td>
</tr>
<tr>
<td>Sense of Control</td>
<td>$r=.17$,</td>
<td>$r=.20$,</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Absorption</td>
<td>$r=.23$,</td>
<td>$r=.20$,</td>
</tr>
<tr>
<td></td>
<td>$p=.07$</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sense of Time</td>
<td>$r=.22$, n.s.</td>
<td>$r=.15$, n.s.</td>
</tr>
<tr>
<td>Self-Consciousness</td>
<td>$r=.02$, n.s.</td>
<td>$r=.11$, n.s.</td>
</tr>
<tr>
<td>Autotelism</td>
<td>$r=.34$,</td>
<td>$r=.35$,</td>
</tr>
<tr>
<td></td>
<td>$p=.006$</td>
<td>$p=.005$</td>
</tr>
</tbody>
</table>

Discussion and conclusion
The aim of the present study was to investigate the relations between individual flow, emotions and perceived socio-cognitive processes involved in computer-supported collaborative problem-solving. First, results show that participants reported having experienced the different states of flow, namely, sense of control, cognitive absorption, distorted perception of time, loss of self-consciousness and autotelic experience. It should be stressed that sense of control is rated as just about average. The reason for this may be that the high interdependence among dyad members makes them feel responsible for both their own- and their partner’s actions. Results also suggest that how the flow dimensions are interconnected in this computer-supported collaborative environment is quite similar to that observed in individual learning situations (Heutte et al., 2016). The two cognitive conditions of flow, sense of control and cognitive absorption, are positively related to each other. There are also relations between flow effects. More specifically, autotelic experience which refers to the affective dimension of flow, is positively related to its two cognitive effects, alteration of sense of time and loss of self-consciousness. These effects are not, however, significantly related to each other. Furthermore, both flow conditions are positively related to autotelic experience. Participants who experience a deep level of involvement in the collaborative process are also more likely to lose track of time. Loss of self-consciousness is the only effect that is not significantly related to flow conditions. It can therefore be assumed that this social dimension of flow (Heutte et al., 2016) would rather depend on the level of well-being experienced during collaboration.
Emotions experienced during the game are significantly related to the conditions of flow and its affective dimension (Question 1). Sense of control and autotelic experience are related to both positive and negative emotions. It is noteworthy that sense of control is more related to positive deactivating emotions and negative activating emotions while autotelic experience is more related to positive activating emotions and negative deactivating emotions. More specifically, an increase in enjoyment and a decrease in boredom. Cognitive absorption is related mainly to positive activating emotions. Participants experience hope when they are intensively focused on what they are working on together. Alteration of the sense of time and loss of self-consciousness are two cognitive dimensions of flow that are not significantly (or marginally) related to the emotions felt by group members during problem-solving.

Interestingly, the individual experience of flow is associated with the ability of participants to accurately assess emotions experienced by their partner during collaboration (Question 2). Furthermore, the flow dimensions associated with modelling the partner’s positive emotions are not the same as those associated with modelling his/her negative emotions. More specifically, alteration in the perception of time is the only dimension of flow related to the accuracy at perceiving the partner’s positive emotions. The mutual modelling of negative emotions is related mainly to sense of control and autotelic experience. The more participants experience control and well-being, the higher their accuracy is at perceiving negative emotions in their partner. Cognitive absorption is significantly related only to modelling negative deactivating emotions, loss of self-consciousness only to modelling negative activating emotions. This means that the more participants are focused on the task, the more they are able to perceive boredom in their partner. It is also easier for them to recognize frustration in their partner when they are less preoccupied with themselves. It is interesting to note that alteration of sense of time and loss of self-consciousness are related to the mutual modelling of emotions but not related to emotions experienced during the task. This suggests that identifying emotions in others would not be necessarily linked to experiencing emotions. It could be rather related to socio-cognitive factors such as the ability to momentarily forget who we are so as to become one with both the activity and others.

There is a relation between flow and the perceived quality of collaborative processes (Question 3). More specifically, all dimensions of flow (except loss of self-consciousness) are positively related to two perceived dimensions of collaboration, i.e. information pooling (gathering information relevant to solve the problem) and transactivity (building on the partner’s ideas). These two categories of processes are recognized as being an important part of the success of collaborative learning (Meier et al., 2007). Therefore, these results mean that participants perceive themselves and their partners as being more socio-cognitively engaged in the task when they experience flow. It is worth noting that flow appears to be related more strongly to participants’ representation of their partner’s engagement in collaboration than to their own perceived engagement. One may thus expect that it would be easier for participants to be aware of their partner’s activity in flow situations.

This study is exploratory. Its limitations lie on the fact that this is a correlational study with only subjective measures collected in a game context (individuals are more likely to experience flow in such a context). However, our motivation was precisely to focus on perceptions (flow, emotions, perceptions of the interaction with the partner), the rationale being that a deeper understanding of the cognitive and affective quality of computer-supported collaborative learning/problem-solving experience is needed to further design environments and tools able to promote positive emotions, engagement and optimal collaborative behaviors (Riva et al., 2012). Results show that the individual experience of flow is related to both the perceived socio-cognitive processes involved during interaction and the mutual modelling of emotions. They also suggest that experiencing a flow state could encourage participants to pay more attention to their partner. Overall, these results support the idea that the flow experience should be considered as a means to improve the quality of collaborative processes and outcomes. A further analysis is needed to investigate the extent to which individual flow is also related to group performance and to objective measures of the quality of collaboration. We are also interested in exploring how to influence the flow experience in CSCL. For instance, the focus could be to manipulate the characteristics of the CSCL environments so as to influence the control and absorption dimensions of flow.

References


Acknowledgements
The experiment described here was designed and carried out with the help of Mireille Bétrancourt (TECFA, University of Geneva), Guillaume Chanel (CISA, University of Geneva) and Thierry Pun (CVML, University of Geneva). It took part of a larger project (EATMINT) on “Affective Computing and Emotion Awareness Tools for Computer-Mediated Collaboration”. EATMINT was a research project of the Swiss National Center of Competence in Research in Affective Science funded by Swiss National Science Foundation.
Epistemic Agency as a Members’ Experience

Danielle Keifert, University of California Los Angeles, keifert@ucla.edu
Christina Krist, University of Illinois at Urbana-Champaign, ckrst@illinois.edu
Déana Scipio, University of California Davis, aeolani@gmail.com
Anna Philips, Tufts University, Anna.Phillips@tufts.edu

Abstract: We present a conceptualization of epistemic agency that centers attention on participants’ experience as a central identifying criterion, including the powered nature of inquiry. This allows us to examine how and by whom the inquiry is defined, navigated, and reified. We draw upon the existing conceptualization of inquiry as a members’ phenomenon to attend to how key epistemic moments—beginnings and endings—allow us to recognize learners’ competence in defining and shaping knowledge products. We take this as an opportunity to sensitize ourselves to issues of power and equity in STEM learning spaces. We argue that viewing epistemic agency as a members’ experience sensitizes researchers and educators to collaboratively decide with learners (rather than for learners) what counts as making progress and identifying paths within and through STEM learning environments.

Introduction

One of the key challenges of the “practice turn” in US science education (Ford & Forman, 2006) is supporting learners’ participation in science and engineering practices, particularly with respect to knowledge building. Informed by research on learning environments that turn over varying degrees of knowledge-building responsibility to students (e.g., Calabrese Barton & Tan, 2010; Engle & Conant, 2002; Scardamalia, 2002), most scholars agree that epistemic agency is a key goal underlying these reforms (Berland et al., 2016; Duschl, 2008; Stroupe, 2014; Windschitl, Thompson, & Braaten, 2008). Existing characterizations of learners participating with epistemic agency often position learners to negotiate paths forward within ongoing inquiry by engaging with the substance of ideas (e.g., arguing that a model of air should include empty space between particles) or navigating investigations once begun (e.g., deciding what roles each group member will take on to conduct an experiment). In contrast, we wish to examine how learners negotiate the context of the inquiry itself: how students and youth negotiate satisfying beginnings and endings to inquiry (e.g., orienting to a new shared puzzle as worthy of investigating, coming to consensus that as explanation is satisfactory). Here we draw on Schegloff’s (1992) context—an understanding of participants’ activity that participants’ themselves show to be relevant during interaction—in this case, how participants show they are orienting to their activity as a moment of inquiry. We also examine negotiations with series of moments in inquiry pathways. We argue that attending to negotiations across moments and pathways sensitizes us to issues of power and equity as we wrestle with how to productively support students’ agentic participation in science knowledge building.

In this paper, we present a conceptualization of epistemic agency that centers our attention on the experiences of participants as a central identifying criterion. This allows us to examine how and by whom the moments of inquiry are defined, navigated, and reified. We begin by drawing upon a framing of inquiry as a members’ phenomenon (Keifert, 2015; Keifert & Stevens, accepted) as a lens for exploring critical moments in inquiry: interactional beginnings and endings. We use this lens to examine two examples of students beginning and ending inquiry on their own terms (over and against their teachers’ plans). We build a conceptualization of epistemic agency a members’ experience (EAME). This conceptualization prepares us to recognize participants’ competence and create space for transformative outcomes. We intend this conceptualization to be drawn upon as a lens not only at the scale of particular interactions (as members’ phenomenon is bound), but also at broader scales of activity across learning experiences. As a result, we see the value of EAME in presenting a view of learning where learners’ experience is the center of educators’ and researchers’ attention for deciding what counts as meaningful participation in scientific work. This re-positions all learners, but particularly those historically disempowered in STEM spaces, to use science as a tool for their own learning and purposes.

An interactional definition of inquiry as a guiding lens for epistemic agency

We begin by briefly presenting inquiry as a members’ phenomenon—IMP (Keifert, 2015, Keifert & Stevens, accepted). IMP helps identify where “productive scientific work” might be happening by recognizing how participants orient to their activity as inquiry. Specifically, we examine beginnings and endings as key moments in any interaction (Jordan & Henderson, 1995). These moments illustrate how participants jointly negotiate what it is that is worth working on together, rather than looking for interactions that fit an a priori description of doing science or inquiry, such as characterizations of argumentation or experiment. Importantly, this lens is not
meant to replace such characterizations or to negate the importance of accountability to the discipline for something to “count” as doing science. Instead, this lens is meant to highlight what may be otherwise overlooked dimensions of the data: it focuses our attention on what participants do that signals productive work to them. We draw upon IMP in the analysis of two moments of inquiry (in a middle school classroom and undergraduate physics course) to examine how students’ experiences with shaping inquiry in beginnings and endings tell us about what counts as meaningful, agentive scientific work to them. This analysis in turn helps us to develop epistemic agency as a members’ experience, a conceptualization we find fruitful at multiple scales.

Inquiry as a Members’ Phenomenon (IMP)

To identify participants’ perspectives about what counts as productive work to them (Sacks, 1967/1992; Stevens, 2010), we draw on a definition of inquiry developed through interactional and ethnographic analyses of video data of everyday interaction (McDermott, Gospodinoff, & Aron, 1978; Jordan & Henderson, 1995). These analyses characterized patterns of 

\begin{itemize}
  \item \textit{beginnings} of joint exploration, as well as the ways that children made progress during inquiry as they drew upon a variety of sensemaking resources. Figure 1 represents these patterns as a prism. The definition of inquiry synthesized from these patterns emphasizes (1) how participants come to begin inquiry as they orient to a shared puzzle, (2) the sensemaking they engage in to make progress in inquiry, and (3) how they end that inquiry as they orient to ending in a manner satisfactory to them.
\end{itemize}

**Figure 1.** The prism of Inquiry as a Members’ Phenomenon (Keifert & Stevens, accepted).

**An IMP illustration: Caroline, Momma, and the Fog**

We present a brief example drawn from this data corpus to illustrate how IMP shows up in interaction and guides our attention to what participants consider to be important and productive work. This example comes from video of Caroline (5y 11m) and Momma at breakfast. Caroline worked to initiate a joint exploration around the fog she noticed outside. Caroline turned her head to the front of the house, then called “Hey Momma” and pointed towards the front of the house, wiggling her hand and adding “Outda window?” in a concerned tone. She then asked, “Why is it so smoky?” After she saw Momma looking out a different window, she repeated this pattern. Caroline’s ‘winding up activity’ (Jordan & Henderson, 1995) carefully coordinated Momma’s attention through talk, gaze, and gesture before she uttered her first question. The range of material and interactional resources she drew upon to coordinate this activity provide indicators that exploring this shared puzzle was important for Caroline.

Ending their joint interaction around the “smoky” also required coordination of material and interactional resources. After walking out the front door to examine the fog more closely, Momma returned and explained to Caroline, “I think it’s just fog...like clouds that are way down low...just blowing by.” Momma oriented to this as a satisfactory explanation for her by turning away to talk to Caroline’s sister. However, after Momma turned away, Caroline asked, “But, has it done that before?” pressing for further explanation. In response, Momma reoriented to their exploration. She explained that while it “doesn’t happen very often here,” “it’s nothing to worry about...[it] happens to be what the weather is today...kinda of low level- low layer of fog”. In response, Caroline turned her attention elsewhere, indicating her satisfaction. In this episode, Caroline and Momma jointly achieved a satisfactory ending only when both oriented to Momma’s explanations as satisfactory. Similar to ‘winding up’ efforts, these ‘winding down’ efforts indicated what kind of explanation was satisfactory for Caroline.

This moment illustrates what we mean by jointly negotiated; it highlights Caroline’s work to achieve shared attention with Momma, their shared exploration of the fog, and efforts to achieve a satisfactory resolution. It also highlights that even very young children like Caroline, age five, take such puzzles seriously as worthy of exploration. This focus on the relevant puzzles and satisfying explanations aligns with recent work emphasizing affect as inherent to inquiry (Jaber & Hammer, 2016; Manz 2015) and work that recognizes children’s diverse sense-making resources (e.g., Nasir, Rosebery, Warren, & Lee, 2006). As an endogenous representation of
Looking for epistemic agency in interactional beginnings and ends

Using IMP to identify moments of productive scientific work helps us decide where to look for the epistemic agency that matters in terms of shaping participants’ experiences of making decisions about beginning and ending knowledge-building interactions. After identifying these interactional beginnings and endings, we then sought to characterize how epistemic agency shows up in those interactions. Rather than viewing epistemic agency as a trait or characteristic of an individual, we draw on Damşa and colleagues (2010) interactional conceptualization of epistemic agency as shared, emergent through “collaborative activities aimed at the creation of shared knowledge objects” (p. 146). In addition, we use Gresalfi et al.’s (2009) operationalization of “agency” as actions (or refrains from acting) that impact the joint activity of a group. When we add “epistemic,” we focus in on actions (or refrains from action) related to knowing and knowledge-building. For example, the choice of using a black pen over a blue pen may be an exercise of agency, but typically not one of epistemic agency. However, if that choice is accompanied by a conversation about how water molecules and air molecules are two different things and should be represented differently, then that choice becomes an exercise of epistemic agency.

We are also explicitly interested in young peoples’ epistemic agency—that is, in understanding how it is that students and youth act in ways that substantially impact collaborative knowledge-building activity in science. This activity often occurs in settings where students are the “novices” and adults are considered the “experts”: K-12 schooling, university classrooms, or programs for youth involvement in professional science labs. Structurally speaking, the bulk of the substantive epistemic decision-making is the responsibility of these expert adults. Therefore, any actions on the part of students and youth that have an impact on their shared knowledge building are the result of a re-negotiation of epistemic responsibility and of power. We examine such instances next.

Epistemic agency in inquiry: Images of some ideal cases

We present two examples illustrating young people’s experiences in substantially negotiating the beginnings and endings of inquiry: their epistemic agency. Examples come from studies conducted in a variety of settings (K-12, university science classrooms) with diverse participants (middle school-young adults). Each vignette represents a substantial re-negotiation of epistemic responsibility and power in ways that allowed learners to be the primary decision-makers at critical points of their inquiry. We use them as vision-casting illustrations: they are not meant to be representative, but instead were intentionally selected to provide concrete cases that expand our notions of what it means to fully embrace reform-based goals of engaging students meaningfully in science practices.

Beginning by identifying and articulating problems

We begin by exploring students orienting to a shared puzzle at the beginnings of inquiry and at pivot points toward new lines of investigation. Determining what is cause for inquiry is a significant epistemic accomplishment. This example comes from a study which identified exemplar moments of students’ seeking coherent, causal explanations of the natural world and to identify themes and patterns across these cases (see Watkins, Hammer, Jaber, Radoff and Phillips, in press; Phillips, Watkins and Hammer (2017). In all cases, students did work to identify, articulate, convince others of inconsistencies or gaps in their understanding, which we refer to as problems or puzzles. In most cases, this process was a primary dynamic, sustained for several minutes and involving multiple participants. In this way, we see this work of orienting towards a shared puzzle as important to the students’ experiences of engaging in inquiry. These shared puzzles become student-created knowledge objects. In no case was this an explicit goal of the instructors; rather the importance of problems as knowledge objects emerged spontaneously within classrooms that fostered epistemic agency.

To illustrate this point, we present part of an episode from a discussion section within a college physics course. During the previous day’s lecture, students discussed a homework question: Does an escalator do more, less, or the same work on you if you walk up or stand still as you go between floors? Most students answered that the escalator does more work. After some discussion about each option, the instructor had given students the answer—the escalator does less work on you when you walk—and moved on. In a recitation section the next day, the TA noticed that students were still discussing the question. She abandoned her previous plans and asked students instead to share their thinking about the question. She abandoned her previous plans and asked students instead to share their thinking about the question. A student, Pat, struggled to articulate her confusion.

Pat: I was saying if we're- Well, if you're walking up the escalator, doesn't the escalator do-well we already know it does less work on you but, like, if for that time, you're still moving up, you're moving up at a constant rate so it's not changing the amount of- wait, ok let me try to articulate this.
TA: Ok. (6.0 second pause)
Pat: Well, when you're walking—no, the escalator isn't doing more work on you while you're walking up for the same argument that we said in class where it was like, your feet is always on the ground, so the same amount of weight is always on the ground, so therefore the escalator is always exerting the same amount of force on you. But if you're jumping for example, then that means that the amount of force the escalator on you— is doing on you, changes (inflected up).
TA: So, if you're jumping, or I would argue, even walking…
Pat: No, if you're jumping on the same step, and you're moving up an escalator, for example,
TA: Oh, if you're jumping on the same step.
Pat: (Overlapping) [Same step]. So, in that case, wouldn't the escalator be doing more force on you, compared to if you're standing still?

Within a brief back and forth with the TA, Pat clarified that her question was about what happens if you jump up and down on the same step as you move up the escalator (Figure 2). Other students then took up this question, driving a new round of discussion around the concepts of force and work. There was still something bothering students, including Pat, about the original homework question. There was still a problem, even though they knew the correct answer to the original question. Pat’s achievement in this moment was not constructing an answer or explanation; it was constructing a new question to address her confusion about the earlier question and answer.

We see students’ epistemic agency, not only in the sense that they had control over the topic they pursued, but also in the sense that they had the authority to decide whether or not there was a puzzle. In this episode, we see Pat’s contribution sparking renewed engagement around the ideas of work and force. In other cases, we see students’ work to articulate their confusion and construct problems sparking new lines of inquiry and sustaining students’ engagement (Phillips, et al 2017; Watkins, et al, in press). We see a complementary response of the educator, allowing Pat and the students the time and power to determine for themselves (at the cost of her lesson plan) a problem worth pursuing. By attending to this beginning, we see how students shifted what counted as open for inquiry, thereby determining for themselves (and with the support of their instructor) a context for inquiry.

Ending inquiries by listening, empathizing, and changing one’s mind
We now explore how shared epistemic agency shapes ending inquiry. This example comes from a study that observed middle school science classrooms to examine students’ participation in science practices developed over three years (Krist, 2016). We present a moment when middle school students oriented to a satisfactory ending to exploration, by their own terms; their experience was one in which they had the power to decide their inquiry was done, and they took on the responsibility to get everyone (including the teacher) on the same page about it.

During the last few days of an earth sciences unit exploring plate tectonics, Mr. M’s 8th grade class was trying to figure out what kind of plate interactions were forming the Andes Mountains in South America. They initially claimed that they were caused by oceanic-continental subduction, when an oceanic plate moves towards and underneath a continental plate, but they noticed two key details that made them question this claim: the trench (a key feature indicative of subduction) was slightly west of the South American coastline; and the South American plate itself was much bigger than South America, meaning it was covered about equally with ocean and continent. These observations led to a several-minutes-long discussion about how they would define whether the South American plate was oceanic or continental.

Before they had reached a decision, Logan asked why it even mattered that they decide which kind of plate it was anyway, eventually proposing that all plates are an alloy-like mixture of basalt and granite:

Logan: Why can’t it just mix together? Like, to make bronze we, we’ve got to mix copper and tin together. Nah, I’m ser—this is on point! […] No, look, to make something else,
you’ve got to mix two things together. You get that, right? So why can’t the plates just like, mix together and then they make like a super-plate?

This idea was highly contentious, and became more heated as Logan drew out his idea on the board and labeled one melting plate as “oceanic” and one melting plate as “continental” (Figure 2). Several members of the class insisted that his labels were a problem, but Logan argued it did not matter what the labels were. There was a lot of cross-talk in response, with several students voicing their opinion that his idea could not work. In the midst of this milieu, Amy, who had been quiet, clarified why Logan’s model did not work:

Amy: I think what I’m trying to say is that like, it wouldn’t make sense because isn’t the oceanic plate like, less dense? I mean--

Logan: No. The oceanic’s more dense.

Multiple speakers: [Unintelligible] why is that? [Several voices] Why is that one there?

Mr. M: Okay, then – I want, I want just Amy and Logan, no one else, for the next minute.

[Several voices; Amy says something inaudible to Logan]

Logan: I get it [erases board]

As Logan erased the board, Mr. M exclaimed, “I don’t get it, what’s this—Wait, ah—ha?!?” He jumped forward from where he had been in the back of the room, sat at a table in the second row, and raised his hand. Logan gestured for him to speak, and Mr. M said, “I don’t get what you just got. Can you say it slowly?” Logan explained that the differences in density in the plates meant that the continental plate could not go under the oceanic plate. Mr. M responded, “Okay, I missed that. So something was mislabeled or something?” While Mr. M struggled to understand why Logan had thrown out his entire model, Courtney summarized for Mr. M that in light of Amy’s idea, only one type of plate could go under the other; so then if the Andes were formed by subduction, the South American plate would have to be continental.

This “ending” of the inquiry was catalyzed by Amy’s simple, and mostly inaudible, comment about plate density. She made this comment after listening to Logan’s idea and others’ critiques of it and pinpointing a piece of information that he was missing that convinced him to change his mind. While Mr. M saw this “resolution” as a decision about the labels on Logan’s model, which perhaps they could continue tweaking, Logan and Courtney saw it as an ending of an exploration around Logan’s idea. This case highlights how students pushed back upon the power structures of schooling in that they move the teacher to their satisfaction despite his expectations/plans. It is important to note, however, that the work they are doing is different than conceptualizations of students’ resistance in the classroom, or intentional non-participation, often in opposition to a teacher’s wishes (e.g., Hand 2010). In this case, the teacher and students were on the same “team,” so to speak; through their negotiation, he was eventually brought on board with their experience of a satisfactory ending.

**Epistemic agency as a members’ experience**

These two examples illustrate epistemic agency as a member’s experience (EAME). Our brief analysis illustrates the interactional accomplishments involved in orienting to beginnings and endings to inquiry and in substantially shaping the nature of those beginnings and endings. This prepares us to recognize the importance of empowering learners to make critical epistemic decisions in key moments such the negotiation of beginnings and endings to inquiry that fundamentally influence subsequent inquiry pathways. In this way, EAME also helps us attend to issues of power and potential transformation across the process of inquiry. We now draw upon EAME to explore learners’ transformation of inquiry paths across the scale weeks engagement in an informal learning environment.
Using Epistemic Agency as a Members’ Experience as a lens to understand youths’ transformation of inquiry paths

Drawing upon epistemic agency as a members’ experience (EAME), we attend now to the impact of histories of power and exclusion in STEM spaces by considering how membership interacts with epistemic agency. Specifically, we draw upon EAME to sensitize us to the contexts learners create (and re-create) for inquiry when they are positioned as epistemic agents throughout the process of inquiring. Considering how people can participate with epistemic agency raises questions not just about the nature of participation but also the goals.

For the purposes of this paper let us consider the concept of making progress through the prism of IMP in Figure 1 (above) by leveraging data at a different scale. Making progress through the prism can be thought of as either potentially benign participation trajectories, or as a set of powered interactions (Esmonde & Booker, 2017). In this case we explore two types of trajectories for youth participating with epistemic agency: a) inclusive and b) transformative (Scipio, in preparation). Inclusive epistemic agency refers to the ways in which youth were able to feel like knowledge creators within existing STEM participation frames (Stroupe, 2014). Transformative epistemic agency refers to the ways in which youth are able to transform science learning spaces to more closely resemble their own ways of knowing and sharing expertise across the longer pathways they navigate through learning experiences. We present here a brief characterization based on extensive ethnographic analysis (Scipio, 2015) to illustrate inclusive and transformative epistemic agency.

Inclusive epistemic agency is built upon work that draws links between everyday science and canonical science practices (Bricker & Bell, 2014; Toomey Zimmerman & Bell, 2014) making it clear that youth can participate in disciplinary ways. While this work repositions youth as knowledge holders and developing experts, these descriptions may also sensitize us to evaluate youth participants’ competence only in relationship to disciplinary practices. Using examples from youth participating in a chemical oceanography out-of-school time (OST) broadening participation program (Scipio, 2015), we can see how powered relationships to disciplinary expectations can shape youth participation trajectories. The OST program was a collaboration between two youth-serving educational programs, a chemical oceanography laboratory, and a learning sciences research laboratory. Youth in the program learned about fish feminization in the local body of water and designed their own research project to leverage the chemical oceanography laboratory’s resources to answer their own questions about water quality. The youth in Scipio’s study participated in the full practices of the collaborating laboratory. They learned how to “make progress” in ways that mapped onto canonical expectations—they collaboratively designed and conducted a research study, they made a poster, and they presented their work at an international conference. As such, their epistemic agency and pathways through the prism were directly related to canonical definitions of STEM participation (similar to the hypothetical students in the classroom in the introduction).

This leads us to ask questions about EAME informed by Megan Bang and Shirin Vossoughi’s questions (2016) about designs in the learning sciences. In particular, can students have experiences in which their decision-making involves “productively disrupting historically powered relations as part of working towards equity and forms of just democracies” (p. 173)? Aiming for such versions of epistemic agency is particularly important for youth from nondominant groups who frequently experience school science spaces as places where settled expectations of normative behaviors and epistemologies are in conflict with their lived experiences and interests (Bang, Warren, Rosebery & Medin, 2012) and also has the potential to disrupt powered relationships in classrooms (Warren & Rosebery, 2011). While inclusive epistemic agency plays an important role in learning, it does not present a way of thinking about youth participating with epistemic agency that allows youth as the members to define the context of inquiry or re-define forms of science participation. Truly redefining epistemic agency as a members’ experience calls for an exploration of the ways that youth participants’ agency transforms STEM spaces (Calabrese Barton, 2001; Calabrese Barton & Tan, 2010). Calabrese Barton (2001) explored this within the context of urban schools where young women redefined what “making progress” in STEM learning spaces could look like. In the OST program, redefining participation led the youth and scientists to collaboratively create new sampling, analysis, and data processing protocols in response to questions posed by youth participants and to transform communication. For example, the team co-created a new sampling kit using glass bottles because the youth wanted to be able to test for chemicals in plastics and the old sampling bottles would have contaminated the water. Youth and scientists collaboratively created a new inquiry pathway—they asked a new question requiring the development of new procedures and materials. The lab then adopted these procedures as part of their standard practice. Thus, youth transformed the practices of this lab as they became members of it. Drawing upon EAME aligned with a critical historical perspective sensitizes us to transformative outcomes.

Discussion and implications

We have presented a conceptualization of epistemic agency that centers our attention on the experiences of participants, especially the powered nature of interactions in which participants (learners and educators) negotiate
exerting epistemic agency is an important first step, and working to create pathways in which they experience that ownership and power is a critical design goal. Many of the existing examples in the literature are cases in which students’ epistemic agency was emergent through a confluence of design factors (e.g., Engle & Conant, 2002); cases of environments explicitly designed to support students’ epistemic agency are rare (see Calabrese Barton & Tan, 2010, Bang et al., 2017 as examples of spaces intentionally designed to support students’ epistemic agency). In part this is because the goals of having learners participate in science practices in ways that position them to engage deeply with the substance of ideas and to take ownership in navigating the investigations that work to build them (Reiser, Novak, & McGill, 2017) are incredibly difficult to realize fully in classrooms. We are drawing upon carefully selected moments to push forward our conceptualizations of epistemic agency, and to remind us that “agency” is not only about providing choice, but that it requires careful decision-making about the nature of those choices. It requires thinking first about how participants would experience the constraints and affordances of those choices, and how that experience might shape future possibilities for decision-making and interaction.

What does this mean for us as researchers and as educators? First, assuming that learners are capable of exerting epistemic agency is an important first step, and working to create pathways in which they experience that ownership and power is a critical design goal. Many of the existing examples in the literature are cases in which students’ epistemic agency was emergent through a confluence of design factors (e.g., Engle & Conant, 2002); cases of environments explicitly designed to support students’ epistemic agency are rare (see Calabrese Barton & Tan, 2010, Bang et al., 2017 as examples of spaces intentionally designed to support students’ epistemic agency). In part this is because the goals of having learners participate in science practices in ways that position them to engage deeply with the substance of ideas and to take ownership in navigating the investigations that work to build them (Reiser, Novak, & McGill, 2017) are incredibly difficult to realize fully in classrooms. We are drawing upon carefully selected moments to push forward our conceptualizations of epistemic agency, and to remind us that “agency” is not only about providing choice, but that it requires careful decision-making about the nature of those choices. It requires thinking first about how participants would experience the constraints and affordances of those choices, and how that experience might shape future possibilities for decision-making and interaction.

We also see how when learners are positioned as competent, we should expect to be unsettled. Attending to issues of power and equity in learning requires that learners be positioned as epistemic agents with the power to transform the context of inquiry, both in regards to the practices and paths. This requires active work on the part of facilitators to support students’ transformative work. It also requires support for educators to learn to navigate with students rather than for students. The examples presented here illustrate how teachers and facilitators had to “let go” of their plans and follow learners. In terms of teacher learning and PD, we need new heuristics and criteria that teachers/facilitators can learn to leverage in supporting students, both as competent epistemic agents, and as transformative epistemic agents who disrupt existing powered relations.

References
of research in education, 30(1), 1-32.


Acknowledgments
We would like to thank Brian Reiser for his feedback on earlier discussions of this work. The projects described in this paper were funded by a National Science Foundation grant to the LIFE Center (#SMA-0835854) and to the University of Washington (#9014483), an Institute of Education Sciences grant to Northwestern (US DOE #R205B080027), a National Science Foundation grant to Northwestern University (#1020316), a Gordon and Betty Moore Foundation grant to Tufts University (GBMF3475), the NAEd/Spencer Dissertation Fellowship Program, the John F. Burlingame Graduate Fellowship in Physics, and the TERC Postdoctoral Fellowship.
Learning Nanoscience Concepts Through a Nanoscale Experience

Polly K. Lai, University of New South Wales, polly.k.lai@gmail.com
Michael J. Jacobson, The University of Sydney, michael.jacobson@sydney.edu.au
Micah Goldwater, The University of Sydney, micah.goldwater@sydney.edu.au

Abstract: Invisible nanoscale phenomena are difficult to learn, as people have no experience of observing nanoparticles and their behaviors in everyday life. To help address this issue, two different treatment conditions were developed: (a) working with two different Agent-Based Models (ABMs) and (b) watching two dynamic visualizations. An empirical study was conducted to compare students in these two groups in terms of reasoning strategies on assessments of declarative and explanatory knowledge as well as solving transfer problems. The findings suggest that the ABM students’ reasoning strategies on the explanatory knowledge tasks and problem-solving activities seemed to have been influenced by the previous computationally embodied learning experience. Nevertheless, the dynamic visualization students seemed to rely on memory retrieval of information from the videos.

Keywords: Agent-Based Model, Dynamic visualization, Nanoscience education, Embodiment of thought

Introduction

Earlier research has argued that students contextualize their science experience into three different mental contexts consisting of imagination, previous experience, and the science investigation itself (Shepardson, Choi, Niyogi, & Charusombat, 2011). In line with this view, it is a big challenge for students to conceptualize and understand the behavior of individual nanoparticles emerges into the properties of matter. This is because their intuition about behaviors of objects is built upon the experience in their daily lives at the human scale instead of the nanoscale (Jones et al., 2013; Magana, Brophy, & Bryan, 2012; Peng, Isaac, & Wilkins, 2012). Thus, the question is how to provide students a nanoscale experience in terms of nanoscience phenomena investigations to enhanced learning of nanoscale concepts? To help facilitate the learning of nanoscience concepts, current research has explored various visualization approaches such 2D static images (Landau, Groscurth, Wright, & Condit, 2009), animated visualizations (Blonder & Dinur, 2011), and 3D virtual environments (Peng et al., 2012). However, research to date on the use of multimedia visualizations in nanoscience classrooms has mainly reported on students’ interests and motivation rather than enhanced deep learning of nanoscience phenomena. Further, there are concerns that merely having students passively watch a visualization would not provide an interactive and cognitively engaging experience to gain a deep understanding of challenging concepts (Chi & Wylie, 2014).

In our previous work, we conducted an initial investigation of comparing students’ learning outcomes of nanoscience concepts after learning with Agent-Based Models (ABMs) and dynamic visualizations respectively (Lai, Jacobson, & Markauskaite, 2016). We found that both treatment groups had a higher declarative knowledge gain; however, the students in the ABM group outperformed the students in the dynamic visualization group on the assessments of explanatory knowledge and transfer problem. To expand our previous work, this study explores the premise that whether learning with ABM provided an opportunity for students to experience an embodiment of thought, which leads to a higher learning gain on the assessments of explanatory knowledge and problem solving. This premise is based on Goldstone and Wilensky’s (2008) argument, which they proposed that the use of “ABMs to teach complex systems topics fits under the wide umbrella of embodied cognition” (p.506). Therefore, according to the wide umbrella of embodied cognition, cognitive processes are influenced and shaped by the body, including the visual perceptual system, body morphology, motor system as well as the body’s interactions with the surrounding world and manipulations of objects in the environment (Barsalou, 2008; Clark, 1998; Gibbs, 2006). This study explores that using ABMs to learn nanoscale phenomena can provide an embodied experience through active hypothesis testing, manipulating individual objects (agents) to increasing an ability, for first-person perspective taking, and eliciting gestures to reason nanoscience concepts afterwards.

Based on the embodied cognition theory, we propose a Framework of Embodiment of Thought to compare and explore the impact of using ABMs and dynamic visualizations to learn nanoscience knowledge: (a) Perspective Taking (Clark, 1998; Soylu, 2016), (b) Causal Inference Making (Gibbs, 2006), (c) Motor-supported Thinking (Mahon & Caramazza, 2008; Schwartz & Holton, 2000). First, research has suggested that ABM learning environments present unique perspective taking challenges to learners (Sengupta & Wilensky, 2009). In ABM learning environments, for instance, students are able to observe an individual agent’s actions such as how
this agent behaves and interacts with its surroundings (a first-person perspective) and accumulates in an aggregate-level representation (a third-person perspective). In contrast, dynamic visualization learning environments provide a storyline animation to represent the scientific concepts only from a third-person perspective. Second, embodied cognition theorists have argued that manipulating objects and exploratory procedures are useful in identifying objects and making goals meaningful (Gibbs, 2006). In other words, when people see a tool, they know what they are going to do with the tool for causal consequences (Gibson, 1986; Schwartz & Holton, 2000; Soylu, 2016). In line with this view, an ABM learning environment allows the learner to run hypothesis testing by setting parameters and exploring procedures and consequences. Thus, the learner knows she is the cause of the individual agent’s behavior, and she also understands the causality directly from her perceptual and embodied experience. However, the information and behavior of dynamic visualizations do not change. The learner has no opportunity to test hypotheses and only passively receives information from the videos. Finally, Schwartz and Holton (2000) argued that mental simulations only mediate coupling of action and imagery in the context of manipulation. Given an ABM environment is allowed learners to manipulate and experience an individual agent behavior and interactions with its surroundings from the first-person perspective, and she also predicts and makes inferences about causality of the agent and the system. After learning with ABMs, she may be influenced by the ABMs and use spontaneous gestures and imagery to form their mental simulations from her past learning experience with the ABMs. Nevertheless, the dynamic visualizations provide a visual representation, learners may form their mental representations through their multimodal perceptions (de Koning & Tabbers, 2011). In short, we argue that the strength of learning with ABMs is how the control of simulations give opportunity for perceptually rich mental models of the nanoscale phenomena, and taking advantage of cognitive capacities in visuospatial thinking, motor simulation, and causal reasoning.

In order to understand why the students in the ABM group outperformed than the students in the dynamic visualization group on the assessments of explanatory knowledge and transfer problems (Lai et al., 2016, April), a study was conducted to compare students’ reasoning strategies on declarative as well as explanatory knowledge and transfer problems after learning from the ABMs and dynamic visualizations respectively.

In line with this, the research questions are:

1. Do students change their reasoning strategies in terms of embodiment of thought from the pre-test to post-test on declarative and explanatory knowledge?
2. Are students’ reasoning strategies grounded in embodiment of thought on the post transfer problems?

This study explored whether students’ reasoning strategies were engaged in the embodiment of thought prior to and after the learning activities. Students’ pre-test and post-test think-aloud protocols were analysed in detail by the three indicators of reasoning strategies in terms of an embodiment of thought framework we propose in this paper. According to the Embodiment of Thought Framework, this study hypothesised that the use of ABMs helps students’ conceptualization development grounded in embodiment thought processes and influence their reasoning strategies afterward, particularly, on the explanatory knowledge and problem solving.

**Methods**

**Research design and procedure**

This study had a mixed method design that consisted of a quasi-experimental intervention, and think-aloud tasks, over one three-hour period (see Table 1). The experimental condition investigated the efficacy of the ABM approach using two Agent-Based Models, which were developed by the first author using NetLogo (Wilensky, 1999). The models provided interactive and manipulative simulations for learning two target nanoscience topics: Nano Gold (see Figure 1 (left)) and Nano Magnetics. In contrast, the comparison condition involved two dynamic visualizations, which were also developed by the first author, for the same topics and activities (see Figure 1 (right)). Finally, think-aloud tasks were used in this study to assess participants’ cognitive processes prior to and after the interventions.

Participants were undergraduate students at Australian universities, who were enrolled in nanoscience programs. Of the twenty-seven participants (average 20.1 years old), seven were females, and 20 were males. Participants received instruction for the think-aloud task and practiced the think-aloud process before the pre-test think-aloud tasks. After completing the pre-test think-aloud tasks, participants were introduced to either the ABM or the dynamic visualization learning environments depending on their group assignment. In the experimental group (14 students), the introduction briefly explained how to use Nano Gold and Nano Magnetics models to run the simulations related to the learning activities. In the comparison group, 13 students were told that they were able to start, stop and rewind the videos as much as they wanted to during the learning activities. Participants then were given two hours to complete learning activities online with the ABMs or the videos individually. Finally,
both groups were administrated the post-test think-aloud tasks, which included all the items on the pre-test think-aloud questions and two extra transfer problems.

Table 1: Study design

<table>
<thead>
<tr>
<th>Activities</th>
<th>Content</th>
<th>Experimental Condition</th>
<th>Comparison Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test think-aloud tasks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Size-Dependent Property in Nano Gold</td>
<td>Gold nanoparticles model</td>
<td>Gold nanoparticles video</td>
</tr>
<tr>
<td>2</td>
<td>Size-Dependent Property in Nano Magnetics</td>
<td>Nano magnetics model</td>
<td>Nano magnetics video</td>
</tr>
<tr>
<td>3</td>
<td>Comparison and Contrast</td>
<td>Compare and contrast two models</td>
<td>Compare and contrast two videos</td>
</tr>
</tbody>
</table>

| Post-test think-aloud tasks |

Figure 1. Nano Gold model (left) and Nano Gold video (right) were used in the study.

Table 2: Pre-test and post-test think-aloud questions

<table>
<thead>
<tr>
<th>Assessment Type</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative nanoscience knowledge tasks</td>
<td>Q1: What is the concept of surface area to volume ratio?</td>
</tr>
<tr>
<td></td>
<td>Q2: Do the individual atoms or molecules of a substance have the same properties as the bulk substance or not?</td>
</tr>
<tr>
<td>Explanatory nanoscience knowledge tasks</td>
<td>Q3: Why does an increase in surface area change the way materials interact with each other on a molecular or atomic level?</td>
</tr>
<tr>
<td></td>
<td>Q4: Why are properties not the same in all three directions (one-, two- and three-dimensional) in a given material?</td>
</tr>
<tr>
<td>Transfer problems (Post-test only)</td>
<td>P1: Imagine you are on a clear sunny day. The sky above us looks bright blue. A few hours later, in the evening, the sunset puts on a brilliant show of reds, pinks and oranges. Why is the sky blue? Why twilight and sunset are in red or orange color? Please write a short essay to answer this question.</td>
</tr>
<tr>
<td></td>
<td>P2: Hard drives and data recording tapes are applications of nanotechnology that depend on magnetic materials. When a magnet is cut into small enough pieces, its magnetic moment becomes increasingly sensitive to the random motion of particles. At a certain point, known as the superparamagnetic limit. Recently, you got a mission from your supervisor to develop a device with maximum storage. Please describe how you will achieve this mission.</td>
</tr>
</tbody>
</table>

Data sources and analysis
Multiple sources of data were gathered prior to, and at the conclusion of the study, including: (a) video/audio recordings of pre- and post-test think-aloud protocols; and (b) video data of participants’ bodily movements during pre- and post-test think-aloud tasks.

The transcripts of participants’ think-aloud protocols were integrated with the video recordings of their bodily movements during the think-aloud tasks, and analyzed in detail by using a coding scheme for reasoning strategies (see Table 3), which was built on the Framework of Embodiment of Thought proposed in this paper. It provided a mechanism to explore whether using ABMs and dynamic visualizations to learn about nanoscience concept might influence participants’ reasoning strategies afterwards, precisely in relation to the target concepts.

Two coders each coded 60% of the subset with a 20% overlap for the coding reliability. The inter-rater reliability was good with the Kappa (κ) = .86, with a significance of \( p < .0005 \). After coding, the percentage of participants, who showed each indicator during the think-aloud tasks, was calculated.
Table 3: Coding scheme for reasoning strategies

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perspective Taking</strong></td>
<td>A first- to third-person perspective</td>
<td>A verbal explanation involves: (a) an agent-level view: illustrating how an individual nanoparticle behave and interact with its neighbours, and (b) an aggregate-level view: illustrating how individual nanoparticles emerge to system properties. You know like the gold that if the particular is smaller, then more space between each other, and the light with different wavelength hits it, resulting in different colours.</td>
</tr>
<tr>
<td></td>
<td>A third-person perspective</td>
<td>A verbal explanation involves: an aggregate-level view: illustrating a system-level phenomenon. Gold nanoparticles at 12 nanometres appear to be red in humane eye.</td>
</tr>
<tr>
<td><strong>Causal Inference Making</strong></td>
<td>Generating causal inferences</td>
<td>A verbal explanation involves idea generating related to causality. The sky is blue because of the size of particles in the atmosphere. The particles in the atmosphere are reflecting and refracting the blue light wavelengths that come from the sun.</td>
</tr>
<tr>
<td>Information retrieval</td>
<td>A verbal explanation mentions prior experience and knowledge</td>
<td>Well, from the video tow. I think the answer is not.</td>
</tr>
<tr>
<td><strong>Motor-supported Thinking</strong></td>
<td>Use of gestures</td>
<td>A verbal explanation supported by gesturing. When the light hits the nanoparticle...</td>
</tr>
<tr>
<td></td>
<td>Use of handwriting/drawing</td>
<td>A verbal explanation supported by handwriting/drawing on a blank paper. This is surface area.</td>
</tr>
</tbody>
</table>

**Results**

**Declarative and explanatory knowledge tasks**

In the ABM group (see Figure 2 (left)), there was an interesting difference between pre-test and post-test think-aloud tasks on Perspective Taking. First, prior to the intervention with ABMs, none of students showed a first- to third-person perspective (an agent-to-aggregate-level view). However, after the intervention with ABMs, 71% of students’ verbal responses demonstrated that they presented their scientific understandings from an agent-based perspective (a first-person perspective) to an aggregate-level perspective (a third-person perspective). Second, 64% of students took a third-person perspective (an aggregate-level view) to explain the concept of structure matter at the nanoscale on the pre-test. In contrast, only 18% of students took a third-person perspective on the post-test. In addition, there was a slight increase in the number of students (64% to 68%) who generated causal inferences from the pre- to post-test think-aloud protocols; on the other hand, the percentage of students on recalling information from previous knowledge to respond to the tasks decreased (100% to 68%). Although none of students took notes to structure their ideas prior to and after the interventions, an increased number of students used gestures to support their thinking (11% to 43%).

Figure 2 (right) shows that most students’ verbal responses reflected a third-person perspective (an aggregate-level view) on both the pre-test and post-test think-aloud tasks (88% and 100%), and none of students used a first- to third-person perspective (an agent-to-aggregate-level view) on either task. Prior to the intervention, although 19% of students drew causal inferences to construct their ideas, after the intervention, all of students
changed to recall previous information for idea generations. Although 8% of students used gestures associated with their ideas, prior to the intervention, there were no gestures at all, after the intervention. In addition, there was a slightly increased number of students taking notes to form their mental representations from the pre- to post-test declarative tasks (12% to 15%).

Figure 2. The percentage of students in the ABM group (left) and students in the Dynamic Visualization group (right) showing the reasoning strategies on pre-test and post-test think-aloud tasks on declarative knowledge.

Figure 3. The percentage of students in the ABM group (left) and students in the Dynamic Visualization group (right) showing the reasoning strategies on pre-test and post-test think-aloud tasks on explanatory knowledge.

For the explanatory knowledge tasks, Figure 3 (left) shows that in the ABM group, prior to the intervention, the verbal responses of most students (82%) showed a third-person perspective (an aggregate-level view) for idea constructions. However, after the intervention, 64% of students provided their explanations in the way, which switched between a first-person perspective (agent-based view) and a third-person perspective (the aggregate-level view) to link the agent-level interactions with emergent outcomes in the ABMs. In addition, 86% of students constructed their ideas on the pre-test explanatory tasks by recalling their previous knowledge. In contrast, on the post-test explanatory tasks, all the students (100%) drew causal inferences for idea constructions. Moreover, there was an increased percentage of students spontaneously using gestures to simulate the nanoscientific phenomena from the pre- to post-test explanatory tasks (32% to 68%). None of the students wrote/drew on the blank paper provided on either task.

In the dynamic visualization group (see Figure 3 (right)), there was an increased number of students providing their explanations from a third-person perspective (an aggregate-level view) from the pre- to post-test explanatory tasks (54% to 88%). None of the students’ verbal responses showed a first- to third-person perspective (an across agent- to aggregate-level view) during their explanations on either task. Additionally, all of the students utilized the similar reasoning strategies to respond to the pre- and post-test explanatory tasks by recalling previous knowledge. It is also noteworthy that fewer students drew casual inferences after the intervention with the videos than prior to the intervention (27% versus 54%). After the intervention with videos, there was a slightly decreased number of students using gestures (19% to 12%), as well as writing/drawing on the blank paper provided for idea constructions (31% to 8%).
Transfer problems
Transfer problems in both groups were conducted after the interventions only. Figure 4 shows notable differences between groups in terms of showing the reasoning strategies during the transfer problems. A Fisher Exact test indicated that the ABM groups had a significantly higher percentage of students demonstrated the solution constructions across an agent-based view to an emergent system view (a first- to third-person perspectives) compared with the dynamic visualization group, $x^2(1, n = 27) = 23.14$, exact $p = .000$, phi = -1.0. In contrast, the percentage of students in the dynamic visualization group was significantly higher than the ABM group in regards to the use of a third-person perspective (an aggregate-level view) on solutions, $x^2(2, n = 27) = 27$, exact $p = .000$, phi = 1.0. Further, all of the students in the ABM group developed solutions based on generating causal inferences, $x^2(2, n = 27) = 14.54$, exact $p = .001$, phi = .73. All the students in the dynamic visualization group searched the solutions from their prior knowledge, $x^2(2, n = 27) = 12.53$, exact $p = .002$, phi = .68. Finally, it is also noteworthy that there was a significant difference between groups regarding using gestures as a support for their problem, $x^2(2, n = 27) = 21.66$, exact $p = .000$, phi = .9. In other words, the ABM group had 85% of students spontaneously using gestures for supporting problem-solving, while the dynamic visualization group only had 8% of students did so. Although 35% of students in the dynamic visualization group wrote on the blank paper provided to support their solutions, none of students in the ABM group used note-taking on the post-transfer problems, $x^2(1, n = 27) = 14.54$, exact $p = .000$, phi = .73.

Discussion and conclusion
The results of this study show that the students in the ABM group during the post-test think-aloud tasks were found to use significantly more of the proposed reasoning strategies in terms of embodiment of thought—(a) Perspective Taking: a first- to third-person perspective, (b) Causal Inference Making: generating causal inferences, and (c) Motor-supported Thinking: use of gestures—than the dynamic visualization group did. For example, the students in the ABM group changed their perspectives from a third-person perspective to a first- to third-person perspective to describe the nanoscale concepts. They did not seem to change their ideas on the post-test tasks by searching their previous knowledge, but rather they started generating causal inferences to conceptualize the nanoscale knowledge in the post-test. This suggests that regarding declarative knowledge, students’ reasoning strategies were slightly influenced by learning with the ABMs. Moreover, the most students changed their reasoning strategies from the pre-test to post-test on explanatory knowledge tasks and transfer problems. For example, during the pre-test, students searched their previous knowledge from a third-person perspective to structure their ideas. However, during the post-test, the ABM students generated causal inferences from a first- to third-person perspective and they were found to spontaneously use gestures as part of their reasoning strategies.

According to recent research in an embodiment framework of cognition, it has been proposed that interacting with tools may change ways an individual thinks and perceives (Gibbs, 2006; Soylu, 2016). In other words, when one manipulates tools, which is usually related to particular goals, this interaction may change an individual’s perceptions and conceptualizations of the environment. The results of this study are consistent with this embodiment view of tools as mediating changes in ways of thinking and perceiving phenomena. For example, the tools used by the ABM group allowed the students to manipulate an individual nanoparticle by setting up the
parameters to observe the individual nanoparticle’s behaviors and interactions (a first-person perspective), and then to explore the consequences of their setting (a third-person perspective). This may have set up opportunities for the students to notice that these parameters settings had a causal impact on the behavior of the nanoparticle (the agent), even though the students did not have direct bodily contact with the nanoparticle. In this way, the ABMs functioned as a learning tool that provided a “virtual” nanoscale learning experience whereby the learner has experiences via their bodily movements with the model, and in turn to see in the behavior of the agents in the model causal consequences of their selections.

This paper proposes that the students in the ABM group had computationally mediated embodied experiences that helped them construct new schema about the nanoscale experience in terms of the target concepts. This in turn allowed them to “run” mental simulations based on these new schema as part of their reasoning strategies when working on the post-test explanatory knowledge tasks and problems. Earlier research has demonstrated a mental simulation recruits the same neural networks involved in action and perception (Gallese & Lakoff, 2005; Jeannerod, 2007), and thus functions as a form of re-enactment of past sensorimotor experience (Borghi & Cimatti, 2010). From this cognitive perspective, it seems that the use of gestures by the ABM group indicated the running of mental simulations during post-test explanatory knowledge and problem solving, and it was the use of the ABMs that provided the “past sensorimotor experience” for students to construct the cognitive representations (i.e., schema) necessary to support these mental simulations.

In contrast, the students watching the videos were only permitted a limited set of interactions with the visualization—albeit with “focus” on selected content information. According to Lai et al. (2016, April), this type of learning activity has been found to foster the acquisition of declarative knowledge but not to be effective for achieving the higher order learning outcomes of explanatory knowledge and knowledge transfer. Most students in the dynamic visualization group had no change from the pre-test to post-test think-aloud tasks in their reasoning strategies in that they took a third-person perspective and recalled previous information to structure their ideas during their assessments. Further, students in the dynamic visualization group demonstrated their reasoning strategies were influenced by the videos in terms of reproductive memory of declarative information and ability to recall images from the videos to draw figures when answer the post-test questions.

In general discussion, this study has found students’ reasoning strategies during the post-test explanatory knowledge tasks and problem solving activities have been influenced by their previous experience of using of ABMs, which suggests that embodied cognition have been a critical aspect of this type of learning experience. This is because students’ mental imagery activates in the ABM group were integral to simulating and activating gestures were highly compatible with an embodiment framework of cognition (Gibbs, 2006; Hostetter & Alibali, 2008). The mental simulations were based on the previous experience of students’ perceptual and motor processes associated with the interactions of the ABM learning environment.

However, current work on embodied cognition and learning design merely focus on physical activity. The design of this body of research was: (a) using physical motion/gestures as a control mechanism for three-dimensional (3D) game-based learning (e.g., virtual environment; Lindgren & Johnson-Glenberg, 2013); or (b) using fingers tracing on paper/iPad as a control mechanism (e.g., Hu, Ginns, & Bobis, 2015); Although the use of physical activity might support enhanced learning, not all learning activities that involve physical activity can be regarded as embodied learning experience.

Moreover, to determine whether physical activities are actual factors in supporting enhanced learning, most research related to action, movement, and gesture studies tends to manipulate physical activities. However, the manipulation of physical activities might complicate research findings. For example, during learning activities, learners’ attention might be disturbed by factors such as remembering gesturing, or not gesturing, rather than learning contents. Another example is that using fingers to trace on paper/iPad might help learners focus on the information they are tracing, and then this concentration leads to learning gains, instead of finger tracing per se. Furthermore, earlier research has claimed that gestures are a natural way of expressing thoughts and spatial information (Hegarty et al., 2005). In line with this view, this study had no manipulation of physical activities. The gestures and handwriting/drawing in this study were students’ strategy choices in a way to express their thoughts spontaneously after learning with the ABMs and dynamic visualizations respectively. In this sense, students moved their hands to simulate the nanoscale phenomena from previous learning experience with the ABMs, and thus, embodied their thought in actions.

In conclusion, the findings of this study are significant in three ways. First, this study has proposed a Framework of Embodiment of Thought with three indicators of reasoning strategy from which to differentiate embodied versus non-embodied learning experiences associated with learning technologies. Second, the findings of this study demonstrated that students in the ABM group spontaneously used gestures to reason the target nanoscale phenomena, which suggests the ABM approach may help students construct new schema and to generate mental simulations as part of their reasoning strategies when problem solving. As this was a small-scale
study, there is limitation to the generalizability of the findings beyond the sample of university students who participated in the study. Overall, it is hoped this study might stimulate further interest in these research areas concerning the nature of innovative learning experiences for understanding difficult scientific knowledge and their underlying theoretical mechanisms.

References


Chinese Character Composition Game for Collaborative Language Learning

Yun Wen, Singapore Centre for Chinese Language, Nanyang Technological University, yun.wen@sccl.sg
Wenli Chen, National Institute of Education, Nanyang Technological University, wenli.chen@nie.edu.sg

Abstract: This paper presents an innovation implementation of whether and how the use of a tabletop game combined with tangible cards help novice Chinese language learners to develop Chinese orthographic knowledge and collaborative learning skill. A positive relation between children’s orthographic awareness and character recognition has been widely acknowledged. Collaboration is a commonly used game design element and a frequently targeted game design element regarding 21st century skill development. Nevertheless, little is known regarding how games should be designed for collaborative Chinese character or vocabulary learning, and how game design may influence student collaborative learning skill development. The results of a quasi-experiment in a primary school indicate that the designed ARC (augmented reality-based Chinese character composition game) is an effective system for improving students’ Chinese character learning performance and collaborative learning quality. The findings provide insights into game-based learning design for young learners and language learners.

Keywords: Game-based language learning; Chinese character learning; collaborative learning; Tangible interfaces

Introduction
Educational game or game-based learning refers to a learning environment where game content and game play help enhance knowledge creation and skills acquisition, and where game activities involve problem solving spaces and challenges that provide learners with a sense of achievement (Qian & Clark, 2016). There has been an increasing interest in effects of digital game-based learning in various domains (Cornillie et al., 2012; Godwin-Jones, 2014; Qian & Clark, 2016). Language learning is no exception. Contextual game-based language learning tends to positively influence language learning (Ericson et al., 2016; Lan, 2015; Wen, 2017). Yet there is a dearth of more empirical evidence concerning how to well design educational game integrating instructional or pedagogical strategies to support active learning (De Grove et al., 2012; Godwin-Jones, 2014).

This study introduces an augmented reality-based Chinese character composition game (ARC), which is designed for Chinese as a second language (L2) learners. It is designed by our research team in line with new Chinese language curriculum of Singapore (The first language in Singapore is English). The study is concerned with Chinese character learning, because Chinese character recognition is a major hurdle for non-native learners and beginning learners. Chinese, as a kind of logographic language, is distinctive from English and other alphabetic languages. The effectiveness of using technological tools on improving learning and teaching of Chinese character has been reported (Zhan & Cheng, 2014). The instantiation of technological use ranges from web-based reading tasks with glossing support to online personal vocabulary learning games or applications. Furthermore, collaboration is a commonly used game design element for engaging player in social interactions, and it is a frequently targeted game design element regarding 21st century skill development. However, little is known regarding how games should be designed for collaborative Chinese character or vocabulary learning, and how games may influence student collaborative learning skill development (Qian & Clark, 2016).

ARC is a tabletop game with tangible paper interfaces designed for enabling collaborative L2 learning in classrooms. It has been using in two Singapore primary schools for one year. In this paper, a quasi-experiment in one Singapore primary school will be presented to demonstrate the effect of the game play on Chinese character learning and collaborative learning. The study aims at not only contributing to the research literature in Chinese character teaching for L2 learners but also shedding light on game-based language learning design for young participants.

Related studies
Coupling physical and virtual objects in character learning
The interactive tabletop environment can provide interesting ways to represent learning content, and thus motivate learners to be engaged in language learning. Multimedia vocabulary learning environments can help
learners construct connections between the verbal and visual representational systems, resulting in an increase in vocabulary knowledge (Chen et al., 2013). Furthermore, interactive tabletops are designed for co-location, multiple user participation, integrating hands-on activates and enabling multiple modes of communication, and their benefits to education have been evidenced as well (Dillenbourg & Evans, 2011).

As for the interactive interface, various arguments have been put forward for why manipulatives (defined as physical objects that can be touched or moved by students to reinforce a concept) may support learning, for instance, providing an additional channel for conveying information, increasing flexibility, empowering students to process and organize information at their own pace, facilitating abstraction and improving memory through physical action (Manches et al., 2010). In the context of language teaching and learning, Corrales (2008) stated that the manipulative can provide opportunities for students to process and organize learning information on their own and at their own pace. Actually, in traditional language teaching, using foreign language flashcards and other manipulatives is a common method to make learning a foreign language fun and exciting. In this study, with the paper interface, the tabletop game not only uses paper as a document with digital capabilities for augmenting its content, but also keeps paper’s intrinsic properties, e.g., tangibility, maneuverability, and flexibility (Prieto et al., 2014).

Character composition for character recognition
Though Chinese character learning is particularly challenging for beginning learners, characters are not random symbols without pattern and regularities. The structure of Chinese characters can be classified by a 3-layer hierarchy: character, component, and stroke. A character is basically constructed by strokes and their combination. Some components are Chinese characters by themselves and these characters are called simple characters or integral characters. Those composed of more than one component are called compound characters. The vast majority of Chinese characters (in the range of 80% to 90%) are compound characters, which are usually composed of a phonetic or semantic component (Shu & Anderson, 1999). Radicals are defined as the meaningful orthographic units that play semantic or phonetic roles in compound characters (Shen & Ke, 2007, p.99). Theoretically, a radical represents the sound of a character or a clue to the meaning of the character (Chen et al., 2013). Radicals have two major features: (1) habitual positions with characters, and (2) function of encoding phonetic information or semantic information of characters (Su & Kim, 2014).

A positive relation between Chinese children’s orthographic awareness and character recognition has been widely acknowledged (Chen et al., 2013). Research also has shown that knowledge of radicals plays an important role in enhancing character learning achievement not only for young school children but also for adult L2 learners (e.g., Shen & Ke, 2007; Su & Kim, 2014). In Gobert et al.’s opinion (2001), it is based on the principle of chunking, in which a chunk refers to a collection of elements having a strong association with one another. Because of chunking, learners can utilize familiar character with a phonetic radical or semantic component to learn and memorize those characters with the identical component but they have not learned yet.

With restricted vocabulary, it may not be easy for beginning learners to realize the importance of character’s radicals. Yet some studies have tested the beginning Chinese learners’ sensitivity to the structures of Chinese characters. They found that the development of semantic radical awareness helped Chinese learners guess the meaning of unknown or unfamiliar characters and revise what has been learned while learning the new (Shen & Ke, 2007). The studies (Ke & Li, 2011; Su, 2010) suggest that the development of orthographic awareness could begin at the first year of study for non-native learners. Indeed, radical-derived character teaching, this instruction approach has become more and more popular in teaching Chinese children characters (Huang, 2008). The effectiveness of this approach was also evidenced by Zhao and Jiang (2002) in their study on investigating 124 non-native learners. It was found that Chinese language use, including summarizing characters with similar pronunciation, meaning or graphic features, appeared to be the most effective in character learning. The findings of Shen’s study (2005) also suggested that systematically introducing radical knowledge to beginning learners can help make sound-shape-meaning connections and greatly facilitate character learning.

In language classrooms, it is sometimes the case that teachers explicitly deliver the radical knowledge to learners. Nevertheless, classroom pedagogy has gradually shifted from knowledge transmission to knowledge construction. Shen & He (2007) compared three types of encoding strategies used in character learning: rote memorization, student self-motivated elaboration, and teacher-guided elaboration. Her findings indicated that elaboration resulted in significantly better retention of sound and meaning of characters than rote memorization. Between student self-motivated elaboration and teacher-guided elaboration, retention of sound and meaning was significantly better with teacher-guided elaboration in study intervals of 20 minutes, but this advantage disappeared at 48 hours recall interval. In a recent study, Shen and Xu (2014) further provided empirical evidence to support the effectiveness of active learning in classroom vocabulary learning for beginning-level
Chinese L2 learners. In other words, student self-directed elaboration can be deemed as an effective approach to learning Chinese characters. Therefore, all the system-based games are designed in radical-derived character learning approach, and students are able to generate and record their own elaboration of the target characters with the support of the ARC system.

**Chinese character composition game: ARC**

In the radical-derived character learning approach, every ARC-based activity is designed to help the student recognize radicals, structures, and compound Chinese characters. A total of six sorts of activities is designed. They are (1) filling characters in a sentence; (2) classifying characters in a paragraph; (3) guessing the character to a riddle; (4) recognizing characters according to the picture; (5) character family; and (6) character connection. Meanwhile, three kinds of paper cards are prepared for students to complete these activities, including structure cards, radical cards, and component cards (see Figure 1). Based on the new Chinese language curriculum of Singapore, the designed ARC game covers approximately 50% compound Chinese characters and over 70% radicals that students need to recognize in primary 1 and primary 2.

To “augment” cards, near-field communication (NFC) readers are used in our system. Like RFID, NFC has advantages of cost-effectiveness and stability of data communication. Every single card is attached to an NFC tag. With this technology, when many cards are on the table at the same time, card information will not be read without mutual interference. Once students tape a card on the NFC reader, its related information will be identified and represented on the iPad screen immediately. In the ARC classroom, students are divided into small groups to play the game together. They are encouraged to exchange cards to complete the activity collaboratively. On the basis of the literature, we assume that students will communicate and exchange ideas a lot with one another. Thus, they would have a good understanding of the target character.

![ARC setting within a small group](image1)

**Figure 1.** ARC setting within a small group

![Self-directed work pace](image2)

**Figure 2.** Self-directed work pace

The activity “filling characters in a sentence”, as an example, in this activity a sentence and its corresponding picture are displayed on iPad screen to provide students the contextual information of the target character. Within a group, students need to discuss the missing Chinese character. After making a consensus, they pick up the structure card for the Chinese character first, and then selected its corresponding radical and compound sequentially. During this entire process, whenever the group cannot make a decision or have no idea about how to continue, they can seek help from the system using the hint card.

After the target Chinese character having been constructed successfully, the group members can use accessory cards to select the exploratory task that they would like to follow. As shown in Figure 3, if a group of students with low language proficiency, they can use the “Pin Yin” card or the “Word & sentence” card to get how the character should be pronounced, or how the character could be used in a concrete scenario. Whilst for a group of students with high language proficiency, they can select those pink color cards to generate their own group artifacts related to the target Chinese character. The card “Let talk”, for example, can be used, when students would like to draw a picture of scenario and verbally make a relevant sentence with the given Chinese character.

**Research design**

A quasi-experimental design was adopted in this study to examine the affordances of the ARC system and their effects on students’ learning performance and competency of collaborative learning. Multiple data sources were triangulated, and the combination of the quantitative and qualitative analysis was used to analyze them.
The participants of this study were grade one students (aged between 7 to 8 years) from a neighborhood primary school in Singapore. Forty-nine students from two classes were involved in our study. During the experimental period, these two classes were taught by the same teacher, Mr A. He is a tech-savvy. He was a computer science engineer, and then came to teach in this school 3 years ago. Both classes received an equivalent amount of instructional time and participated in the similar activities, but the experimental class used ARC game system (N=24), and the control class did not (N=25).

The ARC game designed for primary 1 consists of 4 sections. To keep consistent with school syllabus, our school-based intervention spanned approximately five months, from April 2017 to September 2017. The intervention procedure related to the two classes is detailed in Figure 3. At the beginning, all the students spent 10 minutes to complete a pre-test for testing their Chinese orthographic knowledge. One technical training session (20 minutes) was conducted by our researchers to the experimental class. In the experimental period, each class received an equivalent amount of instructional time and participated in the similar activity design. In both classroom environments, on the basis of the students’ language proficiency, 3 to 4 students were heterogeneously grouped by the teacher and sat together. After the intervention, a 10-minute post-test was delivered to all the participants. Considering the age of students, we conducted a focus group discussion rather than a survey for all the students. Four students per class were randomly selected to participate in the focus group discussion to share with us their attitudes towards playfulness, collaboration, and the game design (20 mins per group).

Additionally, a one-hour professional development session was conducted to Mr A before ARC lesson, to help him being familiar with the system and its design principles. Semi-structured post-interview was administered to get his reflections on teaching and feedback about using ARC.

Data sources and analysis
Data in relation to the learning outcome and processes were collected and triangulated to examine the ARC affordances and effects. Concentrating on the learning outcome, pre-test and post-test were designed to measure students’ Chinese orthographic knowledge. A positive relation between children’s orthographic awareness and Chinese character recognition has been widely acknowledged (Chen et al., 2013). The tests were developed by our research team, referencing a series of tests created by Hung and Fang (2006) and the test created by Chen et al. (2013). We modified the original tests by considering local students’ Chinese language proficiency, and the tests were validated by one primary school master teacher. A total of 40 questions were included in the test with a total score of 40. Gain analysis was used to examine students’ learning performance.

To collect learning process data, during each class, a video-camera was set up at the back of the classroom to record the class process. Meanwhile, two other video cameras were set to record two small groups in each class. The video data of face-to-face group interactions were the main data sources for this study. The video data-based analysis of this study consists of two steps. First, we analyzed and assessed small group collaboration by the rubric developed by Meier et al. (2007). They proposed 9 dimensions to capture the main characteristics of collaboration, and each dimension was rated on a 5-point scale (1=lowest; 5=highest) and the sum of these formed the final collaboration score of each group. In our coding, however, we only included 8 dimensions and excluded the dimension of technical coordination, since it was not suitable for assessing the quality of collaborative processes happened in the control class. Two trained researchers assessed the recorded
data using the rubric for a total of 7 lessons in both the experimental and control classes. The first lesson of the control class was not included because the class grouping had a change after the first lesson. The inter-rater reliability of Cronbach’s Alpha was 0.89.

Second, the data about small group interactions were transcribed verbatim and coded in terms of the concept of collaborative dialogue and the analytical unit of language related episodes (LREs). Collaborative dialogue was proposed by Swain (Swain & Lapkin, 1998; Swain & Deters, 2007) to investigate the act of producing spoken/written language that is the key to students’ understanding of complex linguistic concepts/knowledge. In this study, we described the occurrence of LREs when L2 learners were participated in group activities, in order to examine students’ language learning process in both classes. The qualitative analysis of the video data and field notes helped to explain when the interactions took place or why they did not happen. To make sure the reliability of the data analysis, during the entire coding process, two experienced researchers examined the data, completed the coding independently, and then collaborated and built a consensus on their coding. To make sure the reliability of the data analysis, during the entire coding process, two experienced researchers examined the data, completed the coding independently, and then collaborated and built a consensus on their coding.

Results

Language learning outcome
The learning gain per student was computed by subtracting the post-test score by the pre-test score. Results showed that Mr. A’s ARC class students gained mean=8.61 (n=23, SD=5.92) scores from the learning while the control class student gained mean=-5.43 (n=22, SD=4.12). The independent t-test revealed that the ARC class students made significant improvement compared with the control class students (t=9.19, p<.01). The learning gain of the control class showed negative. This might be caused by the inconsistency of difficulty level of the tests. The pre-test was designed with low difficulty level because it was for students who had just attended primary school. The overall results indicated that students learned Chinese characters with the ARC game made an obvious improvement in orthographic knowledge and awareness.

Collaborative learning process
The histograms in Figure 4 show the mean score of the quality of collaborative learning of each dimension for the two classes. They indicated that groups in the ARC class outperformed than in the control class in most dimensions. Regarding the score of the dimension of time management, however, groups in the control class performed better than in the ARC class. In our class observation, sometimes students in the ARC class spent too much time on the tasks they interested in, especially those exploratory tasks that required students to generate their own group artifacts. We designed the ARC game to provide students with sufficient autonomy that every single group could learn at own pace. Nevertheless, we also noticed that the time management competency of the students at this young age was still weak, let alone in the game-based environment.

No difference was found between the two classes in the score of task division. Based on the class observation, we reasoned that it might be because the problems provided to students were not complicated enough, so they did not have to divide the task to subtasks. Yet it is worth noticing that students in the ARC class did better than the control class in the dimensions of sustainable mutual understanding, dialogue/action management, research consensus, and reciprocal interaction.

Taking time into consideration, we aggregated the numerical scores of all 8 dimensions to obtain a single score to each lesson and displayed them in Figure 5. At the beginning of the intervention, there was no much difference between the two classes. Then the quality of collaborative learning of the ARC class constantly improved. As we observed, with the paper interface, all the students had an equal opportunity to participate, so they were actively engaged. Hence, compared with the control class, the game playing in the ARC class was less dominated by one or two students whose language proficiency were comparatively high. At the first ARC lesson, we observed many students scrambled for playing the game. They continually tapped the cards without discussion, but soon students realized that this way of playing affected their group’s leader board ranking. Students at this age were particularly concerned about badge displayed in games. In the following ARC lessons, we observed that they learned to discuss with others before using the cards. More dialogues about action management could be observed in the ARC class. However, the change in the control class was not obvious.
The retrogression of the control class in the last lesson might be caused by the type of game design. In the control class, students did the tasks without the hints or immediate feedback provided by the system, and thus most often they needed to seek help from the textbook or their teacher when stuck. The activity of the last lesson was about word puzzle that was difficult to find hints from the textbook, so students in the control class were less engaged in this activity. On the contrary, with the hints and contextual animation simulation, students in the ARC class were well engaged.

Besides, though an obvious increment of the collaborative learning score was detected in the ARC class, we had to point out that there was still room for improvement in students’ collaborative learning (see Figure 5). Even at the end, the aggregated score of all the dimensions was only 25.75, which was about 64% of the full score 40. Except for the factors of age and activity design, the characteristics of L2 learning might affect students’ collaborative learning performance as well. As shown in the video data, even though the majority of students in the ARC class were consistently engaged in the activity, not much collaborative dialogue (LREs) took place without teachers’ intervention.

Language-related episodes
The time of each lesson of both the ARC and control classes was approximately 60 minutes. As shown in Table 1, in the ARC class, the average time the teacher provided for student group work was about 44.0 minutes (SD=1.99). Less time of activities was provided for students in the control class and more deviation among the lessons was found (mean=33.94, SD=2.39). According to our classroom observation, the teacher spent less time on introducing the rules of the activities and it was easier for him to provide feedback to students’ work at the class level in the ARC class.

The average number of LREs observed in the ARC class was 5.38 (SD=1.41), and it was 2.50 (SD=1.05) in the control class. Though LREs happened more frequently in the ARC class than the control class, generally speaking, they did not happen frequently during group work process. In other words, even though the ARC game was designed for collaborative learning, the context of L2 Chinese character learning, students talked little about the language they were producing, they seldom questioned their language use or corrected themselves or others. The LREs in the ARC class were more observed when students were selecting radical cards or were generating own group artifacts, for example, in the process of writing the target Chinese character or making oral sentences. An increasing number of the LREs could be found when students were choosing radical cards to compose characters. It suggested that the students became more aware of their physical actions with the manipulatives. Nevertheless, the result was unexpected that the LRE seldom took place when students looking at the pictures or animations providing contextual information about the target character.

Table 1. Activity time distribution and LREs frequency

<table>
<thead>
<tr>
<th>Class</th>
<th>Activity Time</th>
<th>Group</th>
<th>LREs Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L1</td>
<td>L2</td>
<td>L3</td>
</tr>
<tr>
<td>Experimental</td>
<td>41.14</td>
<td>44:35</td>
<td>45:55</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Discussion and conclusion

This study demonstrated an innovation implementation of the use of a tabletop game combined with tangible cards helped novice Chinese language learners to develop Chinese orthographic knowledge and collaborative learning skill. Pedagogically, this study suggested an effective approach to improving L2 learners’ Chinese orthographic knowledge. It provided empirical evidence to support the views of many scholars in the field of game-based language learning, that interactive multimedia and multimode can help to create a contextual learning environment, which can effectively help learners construct knowledge.

Regarding the game-based learning design for young learners and language learners, the results of the study revealed that the tangible interface, together with the badge mechanism, helped to increase the quality of group collaboration. They helped to trigger more coordination during the game play. The use of the tangible cards enabled natural interactions and potentially allowed young students to engage with educational content and collaboration.

From our results, however, it also appeared that in the context of L2 learning, not many in-situ language-related problem solving or discussion emerged in language use. Our initial assumption was that the more language-related discussion or language use would take place when students could collaborate well to complete a task. Yet it was only observed that more LREs regarding the semantic meaning of the radicals emerged in composing Chinese characters. It suggested that students learned to evaluate the radicals selected and thought about ways to compose characters. It revealed that the students were able to reflect on their actions of manipulatives, but they seldom spontaneously discussed the meaning or the context in which the target character could be used. The students showed more concentration on whether the character was correctly composed or not. After that, they paid little attention to the feedback information provided by the game, such as the multimedia contextual information about how to the generated character could be used. Therefore, the future game design could pay more attention to scaffolds that may problematize the subject matter by causing students to pay more attention to critical ideas and connection between new contextual information and existing knowledge of the Chinese character. The type of problematizing scaffold may create opportunities for deeper processing and more productive learning (Reiser, 2004).

For the limitation of the study, the qualitative rating scheme was used to generate a quantitative manure, but it still had the limitation of requiring human judgment. The findings were drawn based on the ARC implementation in a primary one class. Further multiple studies and relevant work will be planned and conducted in the collaborative schools. As the data gets richer, the following studies will further explore the correlation between students’ system-based learning trajectories and their learning outcomes, and place more emphasis on investigating how the manipulatives could be designed and used to encourage students to reflect on their experiences in the game play. The findings will have a broad range of possible implementations for gamed-based L2 learning, beyond Chinese character learning.

References


Learning From Errors – The Effect of Comparison Prompts in Instruction After Problem Solving Settings

Katharina Loibl, University of Education Freiburg, katharina.loibl@ph-freiburg.de
Timo Leuders, University of Education Freiburg, leuders@ph-freiburg.de

Abstract: Students, who engage in problem-solving activities targeting yet to-be-learned concepts, usually generate erroneous or incomplete solution attempts. These erroneous solution attempts can form the basis for acquiring valid target concepts during subsequent instruction. Literature on conceptual change as well as studies on ‘productive failure’ indicate that elaborating on typical errors and comparing these erroneous solution attempts to correct solutions may be crucial for learning in these settings. We compared three conditions in an experimental study: Students of all conditions first engaged in an identical problem-solving activity. Afterwards students worked on elaboration tasks that introduced correct solutions. In this so-called instruction phase, students worked with 1) only correct solutions, 2) correct and typical erroneous solution attempts, 3) correct and typical erroneous solution attempts with prompts to compare these attempts. Posttest results indicate that only students who were prompted to compare the solution attempts significantly benefited from learning with erroneous solution attempts.

Introduction

Imagine the following scenario: students attempt to solve a problem to which they have not yet learnt a solution in school. While struggling with the problem at hand, they generate solution ideas that most likely are incomplete or erroneous (e.g., Kapur & Bielaczyc, 2012). However, these erroneous or incomplete student solutions can form the basis for acquiring valid knowledge during subsequent instruction. An instructional approach that resemble the described scenario is the so-called productive failure approach (Kapur, 2010, 2012): After an initial problem-solving phase, the teacher introduces and explains the correct solution and the underlying concept during the subsequent instruction phase. Multiple studies have shown the effectiveness of productive failure with respect to the acquisition of conceptual knowledge in comparison to an instructional design with reverse order (i.e., instruction followed by problem-solving) (e.g., Kapur, 2010, 2012, 2014; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014; for a review see: Loibl, Roll, & Rummel, 2017). The beneficial effect of the productive failure approach is attributed to the activation of prior knowledge and intuitive ideas in the initial problem-solving phase (e.g., Kapur & Bielaczyc, 2012). Prior knowledge activation, in turn, should help students to integrate new information received during the subsequent instruction phase in their prior knowledge structure (e.g., Sweller, 1988).

However, research showed that the problem-solving activity remains ineffective if followed by an instruction that focusses only on correct solutions (Loibl & Rummel, 2014). In this study, students in the productive failure condition with “problem solving prior to instruction” outperformed their counterparts in the reverse condition “instruction prior to problem solving” with respect to conceptual knowledge only when the instruction built on erroneous student solutions by comparing the erroneous student solutions to the correct solution. Thus, it seems that prior knowledge activation by itself does not fully explain the effectivity of productive failure. In contrast, the finding of this study stresses the importance of building on erroneous solution attempts during the instruction phase.

This finding fits the literature on conceptual change and learning from errors: Comparing erroneous to correct examples focuses students’ attention on the components that differ (e.g., Durkin & Rittle-Johnson, 2012). Thus, comparing erroneous solution attempts to the correct solution focuses students’ attention on aspects that they still have to learn. In other words, students’ knowledge gaps are specified (Loibl & Rummel, 2014). As learning takes place once students realize that they reached an impasse (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003), students need to become aware of their specific knowledge gaps in order to revise their mental model (e.g., Chi, 2000; Vosniadou & Verschaffel, 2004).

In previous studies on productive failure, the teacher (or experimenter) led the comparison between erroneous solution attempts and the correct solution. However, what influences the learning outcomes are the cognitive activities of the learners. Therefore, a relevant open question (that we target in our study) is whether prompting students to engage in such comparisons by themselves can foster learning. Prompts elicit specific learning processes (Renkl 2005), often by fostering self-explanations that learners would not provide by themselves (Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992). Chi and colleagues even found benefits
of prompted self-explanations in comparison to spontaneous self-explanations (Chi, de Leeuw, Chiu, & Lavancher, 1994). Against this background, it is not surprising that prompts have been shown as effective means to foster many kinds of learning activities (e.g., Berthold, Nüikkles, & Renkl, 2007; Devolder, van Braak, & Tondeur, 2012; Schworm & Renkl, 2007). It therefore seems to be a reasonable approach to include prompts during the instruction phase of productive failure in order to support students’ learning processes.

Also, it seems reasonable to argue that triggering these comparison processes is most important for topics with epistemological barriers which require a conceptual change (e.g., Vosniadou & Verschaffel, 2004). One well-studied topic with epistemological barriers is fractions (Prediger, 2008): Knowledge on natural numbers often cannot be applied to fractions, which leads to typical obstacles in understanding relevant aspects of the fraction concept. For instance, typical errors when comparing fractions arise when students focus only on the numerator or the denominator and, thus, produce erroneous graphical and symbolical representations (Eichelmann, Narciss, Schnaubert, & Melis, 2012; McNamara & Shaughnessy, 2011; Stafylidou & Vosniadou, 2004).

As indicated earlier, productive failure has been shown effective for fostering the acquisition of conceptual knowledge. Conceptual knowledge is defined as understanding about underlying principles and structures of a domain (cf. Rittle-Johnson & Alibali, 1999). In other words, conceptual knowledge includes an understanding of the underlying concepts of a solution and it allows reasoning about why and how a procedure works. In addition, there is reason to believe, that comparing incorrect and correct solution attempts during the learning phase fosters the acquisition of so-called “negative knowledge”. Negative knowledge refers to knowing what is not part of a concept and what procedure does not work and why (Oser, Hascher, & Spychiger, 1999). Thus, negative knowledge draws a line between correct and incorrect solution attempts and prevents students from making mistakes (again) that are covered by their negative knowledge. Heemsoth and Heinze (2014) found that reflecting on incorrect examples indeed supported students’ negative knowledge more than reflecting on correct examples. However, in their study this increase of negative knowledge came at the cost of lower conceptual knowledge (in comparison to students that reflected on correct examples) for students with low prior knowledge. Therefore, potentially differential effects on conceptual knowledge and negative knowledge need to be considered.

Research question

Against this background, our main research question is: Does prompting comparisons between incorrect and correct solution attempts after a problem-solving activity foster the acquisition of conceptual knowledge and negative knowledge about fractions? For conceptual knowledge and negative knowledge, we hypothesize that students who are prompted to compare incorrect and correct solution attempts will outperform their counterparts who are not prompted and/or do not work with incorrect solution attempts at all (hypothesis 1).

With regard to the learning process, we investigate, whether students engage in elaborations on errors and comparisons with and without being prompted to do so. We hypothesize that most prompted students will engage in these intended processes, while only few students who are not prompted will do so (hypothesis 2).

Methods

Participants
Participants were 200 fifth-graders from nine classes in Germany. Individual students were randomly assigned to three conditions.

Learning material
The topic fractions had not been covered in class prior to the study. The learning unit of our study covered comparing fractions with graphical and numerical representations. The material relied on the experiences and findings from the KOSIMA project (Kontexte für SI nnstiftendes M Athematiklernen [contexts for meaningful mathematics learning]). The KOSIMA project developed and implemented teaching units for mathematics, including fractions. In our study, we built on core elements of the fractions material from Prediger, Barzel, Hußmann, and Leuders (2013). For more details on the findings on learning fractions with the selected material, see Prediger, Glade, and Schmidt (2011).

In the problem-solving phase, all students were asked to decide which team wins a scoring contest where each player attempts to score a goal once: a team of 5 girls who scored a total of 3 goals or a team of 10 boys who scored a total of 5 goals. It was clarified that each team member only had one attempt as displayed in Figure 1. The first relevant knowledge component is to understand that the absolute number of goals does not
lend itself to a fair comparison as the number of children (i.e., the number of attempts) differs between the groups. Building upon this understanding, the number of goals has to be set in proportion to the number of group members: 3/5 and 5/10. Finally, these fractions with differing denominators need to be compared. This can be done on a graphical level with fraction bars or on a symbolic level by expanding 3/5 with 2.

<table>
<thead>
<tr>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pia’s team (5 girls)</td>
</tr>
<tr>
<td>Ole’s team (10 boys)</td>
</tr>
</tbody>
</table>

Compare the results of the two groups. Who wins – girls or boys? Draw a graphical representation (e.g., boxes or bars) to explain your decision. Label your representation with fractions.

**Figure 1.** Translated task in the problem-solving phase.

In the instruction phase, all students received two canonical solutions marked as correct. The students in the error condition and in the prompt condition additionally received two erroneous solution attempts (focus on number of goals only; comparing number of goals and number of group members separately) marked as incorrect. Figure 2 provides an example of an erroneous solution attempt.

**Figure 2.** Translated example of an erroneous solution attempt provided in the instruction phase.

All students received a written sheet that specified their tasks. This sheet included different prompts for the different conditions: Students in the control condition and in the error condition received the following general prompt: “Have a look at the solution ideas. Was the comparison fair? Who wins?” The sheet did not
specify whether the students should elaborate on the solution attempts one by one or overall. Students in the prompt condition received more specific prompts to compare two solution attempts (a correct and an incorrect attempt) at a time, to identify the error in the erroneous solution attempt, and to explain what is better in the correct solution (e.g., “Compare the solution ideas of Till and Ole. What did Till do wrong? What did Ole do better than Till?”). Finally, all students were asked what they need to consider in general in order to compare fairly when the number of group members differs (e.g., when 3 kids compete against 9 kids).

**Measurements**

Prior to the study, we administered a prior knowledge test, a test regarding students’ mathematical operations sense, and demographical data. At the end of the study, we administered a posttest testing for conceptual knowledge and negative knowledge.

The test for prior knowledge covered items related to the learning unit (isomorphic to the conceptual knowledge posttest items, see below). The prior knowledge test did not include any items targeting negative knowledge. Including negative knowledge items (cf. posttest items below) in the pretest would have altered the exploration phase by providing solution ideas and would have interfered with our manipulation by providing erroneous solution attempts to all students.

The mathematical operations sense was tested with items adopted from Lernstand 5, a central comparative test (Schulz, Leuders, & Rangel, 2017). The test included 10 items such as “Pia’s dad is 48 years old. He is 6-times as old as Pia. How old is Pia? Please write down your calculation (not the result).”

Finally, students filled in a short questionnaire on demographical data asking for gender, age, first language, and mathematics score of the last school report.

The posttest included items on conceptual knowledge and items on negative knowledge. The items on conceptual knowledge introduced problems similar to the one in the problem-solving activity and asked students to explain their reasoning. The negative knowledge items presented student solutions to similar problems and asked students to identify whether these solution attempts are correct or incorrect and to explain their reasoning. The maximum was 7 points for conceptual knowledge (6 points for correctly identifying the biggest fraction and reasoning in each case plus 1 point for general reasoning) and 12 points for negative knowledge (for identifying six solution attempts as correct or incorrect and reasoning in each case).

Regarding our process measure, we analyzed students’ answers to the elaboration task in the instruction phase. We coded a) whether students explicitly referred to at least one of the erroneous solution attempt and b) whether they explicitly identified the error. Students from the control condition were excluded from this analysis because they did not receive erroneous solution attempts during the instruction phase and, thus, were not able to refer to them in this phase.

**Procedure**

In order to keep the instruction equal across conditions, all classes were taught by trained teachers, who did not know our hypotheses, but were aware of the differences in the learning material. The study started with an introductory session for all conditions to guarantee a basic understanding of fractions (e.g., division of 4 pizzas for 6 children without remainder). This session was followed by a prior knowledge test.

During the main session of the study, all students first worked for 25 minutes individually on a problem on comparing fractions with unequal denominators (problem-solving phase). The previous introductory session covered neither comparing fractions nor expanding fractions (in order to obtain a common denominator). Thus, the target concept was unknown to the students. The problem-solving activity was identical across conditions. Students handed in their solution attempts at the end of this phase.

Afterwards students worked for 20 minutes on elaboration tasks which introduced canonical solutions (instruction phase). This phase differed according to condition as displayed in Table 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Correct solutions</th>
<th>Erroneous solution attempts</th>
<th>Comparison prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control condition</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Error condition</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prompt condition</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

In the control condition, student received two canonical solutions to the problem. They were asked to elaborate on the solution attempts and to explain how to solve such a problem in general. In the error condition,
students additionally received two typical erroneous solution attempts marked as such. The elaboration tasks were the same as in the control condition. Thus, in the error condition students were able to draw comparisons between correct and incorrect solution attempts on their own, but they were not explicitly prompted to do so. In the prompt condition, students received the correct and erroneous solution attempts and were additional prompted to compare two solution attempts at a time before answering the elaboration task. Students of all conditions were asked to write down their answers to the elaboration task. There answers (as well as the correct and erroneous solution attempts) were collected at the end of the instruction phase. Students of all conditions worked individually in the same classroom. Time on task was held constant across condition.

Afterwards all students completed a posttest targeting conceptual knowledge and negative knowledge. The posttest took 25 minutes. All students finished the test in time.

**Results**

**Prior knowledge**

The conditions did not differ in their prior knowledge on comparing fractions \(F(2,194) = 0.234, p = .79\), regarding their mathematical operation sense \(F(2,193) = 0.153, p = .86\), nor regarding their general math score from the last school report \(F(2,151) = 0.330, p = .72\). Table 2 displays an overview of these control variables.

**Learning outcomes**

About 20% of the posttest data was coded by a second independent rater. Interrater-reliability (ICC) was high for both scales, conceptual knowledge \((IC_{2,1} = .993)\) and negative knowledge \((IC_{2,1} = .954)\). Only the 196 students with complete datasets were included in the following analyses. Table 3 presents the mean test scores for the three conditions.

Table 2: Means (SD) of the control variables.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prior knowledge on comparing fractions (max. 7 points)</th>
<th>Mathematical operation sense (max. 10 points)</th>
<th>General math score (1 to 6, 1 is the best score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control condition</td>
<td>1.58 (2.01)</td>
<td>5.55 (2.14)</td>
<td>2.15 (1.17)</td>
</tr>
<tr>
<td>Error condition</td>
<td>1.36 (1.67)</td>
<td>5.75 (2.12)</td>
<td>2.18 (1.19)</td>
</tr>
<tr>
<td>Prompt condition</td>
<td>1.49 (1.93)</td>
<td>5.66 (2.00)</td>
<td>2.31 (1.05)</td>
</tr>
</tbody>
</table>

Table 3: Means (SD) of the posttest scores.

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Conceptual knowledge (max. 7 points)</th>
<th>Negative knowledge (max. 12 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control condition</td>
<td>64</td>
<td>3.03 (2.56)</td>
<td>4.27 (2.72)</td>
</tr>
<tr>
<td>Error condition</td>
<td>64</td>
<td>3.33 (2.22)</td>
<td>4.75 (2.67)</td>
</tr>
<tr>
<td>Prompt condition</td>
<td>68</td>
<td>3.81 (2.47)</td>
<td>5.47 (2.90)</td>
</tr>
</tbody>
</table>

Prior knowledge on comparing fractions (conceptual knowledge: \(r = .491, p = .000\), negative knowledge: \(r = .220, p = .002\)) and the mathematical operation sense (conceptual knowledge: \(r = .445, p = .000\), negative knowledge: \(r = .291, p = .000\)) significantly correlated with the posttest scores and were therethrough included as covariates. ANCOVAs with prior knowledge and mathematical operation sense as covariates revealed (marginal) significant effects for conceptual knowledge \(F(2,191) = 2.70, p = .07, \eta^2_p = .03\) and for negative knowledge \(F(2,191) = 3.46, p = .03, \eta^2_p = .04\). Posthoc comparisons (LSD) revealed significant differences between the prompt condition and the control condition, both for conceptual knowledge \((p = .02)\) and for negative knowledge \((p = .01)\), favoring the prompt condition. Differences between the error condition and the control condition (conceptual knowledge: \(p = .34\), negative knowledge: \(p = .31\)) and between the error condition and the prompt condition (conceptual knowledge: \(p = .19\), negative knowledge: \(p = .12\)) were not significant. Taken together, our results partially support hypothesis 1: The prompt condition significantly outperformed the control condition on both scales, but there was no significant effect compared to the error condition.
Process data
Our process analyses yielded insights on whether students in the error condition (without being prompted) and students in the prompted condition did attend to the erroneous solution attempts. From the 65 students in the error condition only 6 referred to an erroneous solution attempt and from these 6 only 2 identified an error. In contrast, in the prompt condition 67 of 68 students referred to the erroneous solution attempts and 56 of them identified at least one error. Thus, our results descriptively support hypothesis 2.

Discussion
The research presented here focuses on the instructional approach productive failure (cf. Kapur & Bielaczyc, 2012; Loibl et al., 2016). Productive failure combines the generation of divergent solution attempts in a problem-solving phase with a subsequent instruction phase. Our study focused on the processes in the instruction phase. More precisely, we investigated whether elaborating on erroneous solution attempts and comparing them to correct solutions after problem solving fosters learning fractions. To answer this question, we compared three conditions: Students of all experimental conditions first engaged in an identical exploration activity. In the subsequent instruction phase they worked on an elaboration task that introduced either a) only correct solutions (control condition), or b) correct solutions and typical erroneous solution attempts without prompts (error condition) or c) correct solutions and typical erroneous solution attempts with prompts to compare these solution attempts (prompt condition).

Our results indicate that including erroneous solution attempts is beneficial for learning (both conceptual knowledge and negative knowledge), if students are explicitly prompted to compare them to correct solutions (prompt condition): At posttest, students in the prompt condition significantly outperformed students in the control condition who studied correct solutions only.

In contrast to our results, Heemsoth and Heinze (2014) found beneficial effects of elaborating on errors on negative knowledge and not on conceptual knowledge. These divergent findings might be explained by the fact that the study by Heemsoth and Heinze compared two slightly different conditions: elaboration on correct examples only (cf. our control condition) versus elaboration on incorrect examples only. None of the conditions engaged in comparisons between correct and incorrect examples (cf. our prompt condition). Thus, our results indicate that including erroneous solution attempts is only beneficial for the acquisition of conceptual knowledge, if students are explicitly prompted to compare them to correct solutions.

In our study, students who received the erroneous solutions without comparison prompts (error condition) did not differ significantly from the other conditions. To some extent, this result supports the notion that confrontation with errors alone does not unfold the full potential of learning from errors (cf. Große & Renkl, 2007; Heemsoth & Heinze, 2014). As the error condition and the control condition did not differ significantly on any tests, it is of interest whether students in the error condition actually attended to the erroneous solution attempts. While the implicit cognitive activities naturally remain unrevealed, our process data shows that almost no student in the error condition explicitly elaborated on the errors. In other words, the students in the error condition did not take advantage of the opportunity to elaborate on errors (cf. hypothesis 2) and their learning process therefore resembles the process of students in the control condition (at least on the surface level). This finding underlines the importance of differentiating between the intended learning process and the process that actually takes place (cf. Dillenbourg, Baker, Blaye, & O'Malley, 1996). In contrast to the error condition, almost all of the prompted students explicitly referred to the erroneous solutions and identified the errors while working on the elaboration task. Thus, in line with our hypothesis 2 our prompts triggered the intended processes that, in turn, seem to be beneficial for learning. However, the link between the process and learning is rather speculative, because the prompt condition did not significantly outperform the error condition at posttest. Future research should investigate in more details the different processes and their relation to learning.

While attempting to keep materials as parallel as possible, we have to acknowledge that the number of tasks and prompts differed between conditions. Despite this difference, time on task was held constant across condition. This was possible as the conditions with fewer prompts (i.e., the error condition and the control condition) received broad prompts, while the condition with more prompts (i.e., the prompt condition) received very specific prompts. Our hypotheses focus the type of prompts. However, the number of prompts also may have affected the learning outcomes. Future research may detangle the effects of number of prompts and type of prompts.

Inspired by the in vivo research paradigm advocated of the Pittsburgh Science of Learning Center (Koedinger, Corbett, & Perfetti, 2012), we conducted our study in schools during regular mathematics lessons. Nevertheless, we kept the internal validity high by highly standardizing our procedure. However, this highly controlled study design comes at a cost: Except of the experimental manipulation, we kept the instruction phase
equal for all students. That is, all students received the very same correct and incorrect solution attempts, independent from what they themselves produced during the problem-solving phase. Potentially, a flexible instruction that reacts to individual solution attempts of students in a more specific and adaptive way could be even more effective. We aim to test this hypothesis in a future study by using a computer-based adaptive environment.

References
Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. Learning and Instruction, 34, 74-85.


**Acknowledgments**

This research was supported by the Deutsche Forschungsgemeinschaft, DFG (LO 2196/1-1). We would like to thank the participating schools and students. Thanks to our student assistants for their help in collecting and coding the data.
Identifying Reflective and Non-Reflective Group Consensus Strategies for Evidence-Based Scientific Argumentation

Susan A. Yoon, University of Pennsylvania, yoonsa@upenn.edu
Miyoung Park, University of Pennsylvania, parkmi@gse.upenn.edu
Emma Anderson, Massachusetts Institute of Technology, eanderso@mit.edu

Abstract: This study examines how students come to agreement on evidentiary data they collectively use to support scientific claims or explanations. Student groups were videotaped participating in high school biology units that were scaffolded in inquiry-based experimentation and argumentation with complex systems computer simulations. Interactions around the argumentation prompts revealed extraneous/non-reflective and generative/reflective strategies for collective evidence use. We hypothesize that access to recorded archived data along with access to dynamic just-in-time simulations promoted more generative/reflective strategies. Preliminary data on students’ written responses reveal stronger written argumentation responses when students engaged in generative/reflective strategies compared to when students used extraneous/non-reflective strategies. Where extraneous/non-reflective strategies occur, we suggest that teachers be made aware of the tendency toward this kind of group decision making and that greater emphasis be placed on optimal evidence use through active modeling.

Introduction

A central scientific practice that has shown great promise and great challenge in supporting student learning is scientific argumentation. Argumentation enhances students’ conceptual understanding of scientific phenomena, while also strengthening skills in scientific reasoning (Osborne, 2010). However, researchers have found that building argumentation skills and explanations based on evidence is not easily learned (Kuhn, 2010). Some research has shown that an emphasis on constructing explanations can prevent students from connecting evidence to claims (Kuhn et al., 2013; Kuhn & Katz, 2009). Due to a social aspect that can influence the argumentation process while defending claims, students may use persuasive discourse rather than objective appeals to evidence (Berland & Reiser, 2009; 2011). Similar non-content goals, such as task completion, have been shown to inhibit students’ attention to meaningful data interpretation (Ryu & Sandoval, 2015). But considering the importance placed on group collaborative argumentation, surprisingly few studies have focused on understanding the mechanisms in evidence use. Examining how students come to agreement on evidentiary data can provide insight into developing instructional supports that can overcome the tendency toward persuasion. The research questions underpinning our investigation are: (1) What strategies do students working in groups use to come to agreement on the evidence they use to support claims? (2) What is the predominant strategy used? (3) How can more reflective and evidence-based strategies be supported? (4) What do students’ written argumentation responses tell us about the use of reflective and evidence-based group consensus strategies?

Background

Manz (2016) summarizes the literature illustrating the struggle that K12 students have in constructing and using evidence, which includes not attending to relevant data, ignoring data that are inconsistent with their understanding, and failing to consider sources of uncertainty and error. However, as she points out, scientific knowledge is socially constructed, which underpins the central concern about the lack of accurate use of evidence in argument through persuasion that Kuhn et al. (2013) and others have raised. Thus, it is important to examine how groups come to consensus on their claims or explanations and how evidence is used in the process. Here we point to previous work that we have done on non-reflective and reflective decision-making strategies that can be useful in framing our understanding. Yoon (2011) suggests that non-reflective strategies promote extraneous processing (Stull & Mayer, 2007), in which learners make decisions in ways that are unrelated to the instructional goal. Reflective strategies promote generative processing, in which learners engage in deeper cognitive work and attend to relevant information. We anchor our analyses in this framing to better understand the consensual processes students use in performing argumentation tasks while using simulations of complex scientific phenomena.

Methods
Context and participants

This study is part of a larger project focused on developing curriculum and instruction to support learning about complex systems in high school biology conducted 2010–2015. The broader intervention involved implementation of units on the topics of genetics, enzymes, diffusion, evolution, ecology, and modeling. The 3-day units directed students to perform multiple experiments using different initial conditions such as varying population sizes, energy characteristics, and varying numbers of interacting simulation components. Students recorded their observations and responded to tasks based on their data collection such as constructing graphs. Each unit included prompts for students to come to consensus using the argumentation structure of claims, evidence, and reasoning (Novak et al., 2009). Figure 1 depicts an example of an argumentation prompt.

![Figure 1. Argumentation prompt in the evolution unit.](image)

In this study, we sampled 70 (35 females, 35 males) out of 361 students we worked with in the larger study.

Data sources and analysis

We conducted analyses on 33 videotaped interactions of student groups (mainly pairs). Coding was completed through a method of interaction analysis (IA). This involves analysis of video clips by a group of researchers to examine the details of social interactions (Jordan & Henderson, 1995). Codes and categories in IA are meant to emerge from multiple replays of video data and negotiation of coder interpretations. This is meant to mitigate individual researcher bias and enables mutual construction of meaning. While performing IA in this study, the research group (authors), referred to previous research findings that indicated content and non-content decision making processes. Codes emerged as either extraneous or generative in terms of content goals. Extraneous strategies showed non-reflective behavior where students did not engage with evidence previously collected. Generative strategies showed reflective behavior where students actively engaged with evidence in different ways. Codes, definitions and exemplars are presented in Table 1. We also present examples of transcribed footage that provide deeper insight into how students discussed the argumentation prompt and worked with evidence to put forth a particular claim.

To understand how students’ claim selection strategies impacted their ability to construct an argument, we conducted a preliminary analysis that examined argumentation responses from their unit packet worksheets. We present two cases of students’ CER (claim, evidence, reasoning) responses for two units that showed promising evidence of the relationship between generative/reflective argumentation strategies and correct scientific explanations. In the first case on the topic of sugar transport, students were asked to construct an argument from the following general prompt, “What is responsible for the spreading of the molecules you just observed?” In the second case with the unit on enzymes, students were provided a choice of claims to select from. The following directions were given on the worksheet:

*How do the enzyme and starch (substrate) come together to interact? Discuss the following possibilities with your group, choose the ONE claim (either A, B, or C) you think is most likely, and write down your group’s evidence and reasoning for that choice. Run the Experiment 2 simulation as many times as necessary to establish your claim. Claim A: Enzymes are drawn to substrates, like a hungry traveler without a map, in a new town, who smells pizza from a distance and heads towards the scent. Claim B: Enzymes find substrates, like a hungry traveler without a map, in a new town, actively looking for thick crust pizza as they walk down the street. Claim C: Enzymes find substrates, like a traveler without a map, in a new town, wandering the streets in no particular direction, until they bump into a pizza place. Pizza happens to be the only food they like, so they go inside to eat.*
<table>
<thead>
<tr>
<th>Category and Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extraneous/Non-Reflective Group Argumentation Strategies</strong></td>
<td></td>
</tr>
<tr>
<td>Follow the Leader</td>
<td>A student asks the other students in her group if they know the answer. The other students don’t know the answer and do not show any initiative to find out the answer. First student reruns the experiment but is the only one looking at the screen and then proceeds to tell the other three students the answer to the argumentation question. The other three students simply write down what she said. (Teacher ID 1, Group 1-2, Diffusion Unit, 2/11/14)</td>
</tr>
<tr>
<td>Work Independently</td>
<td>Two students work on their own worksheets. One student in the pair says “it’s claim A.” He waits for the other group member to catch up who says “so, its claim A?” and the two independently complete their answers to the argumentation question. (Teacher ID 9, Group 3, Genetics Unit, 12/10/2013)</td>
</tr>
<tr>
<td>Teacher Facilitated: Non-Reflective</td>
<td>Teacher encourages the pair of students to “Play [the simulation] again and see what you see.” Student B sits quietly and waits until student A has figured out the answer. Student B does not engage with the teacher’s questions. (Teacher ID 9, Group 2 Genetics Unit, 12/10/2013)</td>
</tr>
<tr>
<td><strong>Generative/Reflective Group Argumentation Strategies</strong></td>
<td></td>
</tr>
<tr>
<td>Use of Already Collected Evidence</td>
<td>Two students discuss the argumentation prompt. During their exchange the students point to the graph on the screen to confirm their understanding. (Teacher ID 9, Group 1 Genetics Unit, 12/10/2013).</td>
</tr>
<tr>
<td>Collect New Evidence</td>
<td>Before selecting their group’s claim, they rerun the simulation and discuss how the enzymes are acting. Eventually the students agree that the enzymes show random motion based on what they collectively see on the screen. (Teacher ID 1, Group 1, Enzymes Unit, 11/22/2013)</td>
</tr>
<tr>
<td>Teacher Facilitation: Reflective</td>
<td>Teacher encourages the pair of students to “Play [the simulation] again and see what you see.” Student A replays the simulation and the teacher directs the student’s focus to the relevant information. (Teacher ID 9, Group 2 Genetics Unit, 12/10/2013)</td>
</tr>
</tbody>
</table>

**Results**

In total, 38 codes were assigned to the video interactions of students in their argumentation groups of which 5 were double coded. Figure 2 shows the distribution of strategies across codes. In general, student groups used more generative/reflective strategies. Of those codes, the highest number, 14 (38%), was found in the Use of Already Collected Evidence category. We believe that the curricular layout with scaffolded experiments and argumentation prompts enabled students to refer back to archived and easily retrievable information. A relatively high number of codes, 7 (18%), also emerged in the Collect New Evidence category. Here, it is likely that the use of simulations enabled the dynamic, just-in-time feedback to occur where group members could watch the same
screen and agree on the same evidence. However, a large number of codes, 9 (24%), was found in the *Follow the Leader* category, which we hypothesize may have negative consequences on some students’ learning.

In Figures 3 and 4, we show excerpts and commentary from discussions as students engaged in the two highest categories of generative/reflective strategies to come to consensus on the argumentation prompt (see endnote for discourse analytic conventions). In the first excerpt, shown in Figure 3, two students, B (boy on left) and G (girl on right) converse with each other and use previously collected data to understand how gene regulation occurs.

**Discourse Interaction Excerpt**

<table>
<thead>
<tr>
<th>Discourse Interaction Excerpt</th>
<th>Commentary</th>
</tr>
</thead>
<tbody>
<tr>
<td>B (boy on left) and G (girl on right)</td>
<td>The group starts their discussion off by first figuring out how a home heating system works (lines 1-8).</td>
</tr>
<tr>
<td>1. B: ((Reads question aloud.)) Ummm…Claim A, a gene regulatory system is like a home heating system that is controlled by a thermostat. Is not like a…((Laughs))…ummm ok</td>
<td>In line 9, B asks the important question of whether gene regulation works in the same way that a home heating system (that they described) works. In response, G asserts that gene regulation is like a home heating system, but has difficulty explaining why when pressed by B (lines 11-12).</td>
</tr>
<tr>
<td>2. G: What do you mean?</td>
<td></td>
</tr>
<tr>
<td>3. B: A gene regulatory system is like a home heating system…so you turn the heat up…and then it like goes up?</td>
<td></td>
</tr>
<tr>
<td>4. G: Oh no it’s like the one when you, it does it itself…kind of, no wait. What?</td>
<td></td>
</tr>
<tr>
<td>5. B: I dunno what you’re talking about.</td>
<td></td>
</tr>
<tr>
<td>6. G: Cause the heat at my house, when it gets like, too hot, it turns off. But if it’s too cold, it turns on, right? Right?</td>
<td></td>
</tr>
<tr>
<td>7. B: Ok. So, you’re like…ok, I think that’s how heating systems work.</td>
<td></td>
</tr>
<tr>
<td>8. G: Ok, so it is like that, right?</td>
<td></td>
</tr>
<tr>
<td>9. B: Well is that like that, though?</td>
<td></td>
</tr>
<tr>
<td>10. G: Yeah…</td>
<td></td>
</tr>
</tbody>
</table>

(2)

11. B: You sure? So when it makes, so when there’s a lot…
12. G: So when there’s a lot of lactose…
13. B: So when it’s up, it goes… ((Gestures with hands going up and then down))

(19)

14. B: Ok, so when it makes mRNA, right?
15. G: Yeah…yeah
16. B: Then enzymes get made to eat the lactose (Gesture with both hands where one hand is engulfing the other hand)...but the lactose pulls away the stuff.
17. G: Yeah, yeah so it is...yeah! And mRNA eats the lactose.
18. B: Yeah.
19. G: So it is.
21. G: It is! Cause when there’s too much, it sends something out to stop it. That’s why you don’t have hair on your liver cells.
22. B: Okay.
23. G: Right? (Laughed)
24. B: I guess? (Laughed)
25. G: Claim A.
26. B: So our evidence...

(5)

27. B: ...for this
28. G: It tries to like, get a medium.
29. B: Is that our evidence, or our reasoning?
30. G: That’s my reasoning.
31. B: So our evidence is that...the enzymes eat the lactose?
32. G: The evidence is that like...when the RNA polymerase goes in the DNA...(Points her pen to the simulation that had been previously run and paused)
33. B: Oh it’s like feedback! Like positive feedback...or negative feedback.
34. G: Yeah, so it’s...ummmm...which one is it? I was thinking earlier, but I forgot the name.
35. B: I forget the difference...like, I know what the difference is...but I forget the...like which one’s which.
36. G: Negative’s when you add something. I mean, negative’s when you take something away and positive’s when you add something.
37. B: So, positive, it’ll just like go up...right? So then negative...it has to be negative then.

As the conversation ensues, B starts to explain how “enzymes get made to eat the lactose” (line 16) and G agrees (line 17).

At this point, it seems that G is convinced that, in gene regulation, “it sends something out to stop it,” (line 21) which is similar to a home heating system’s feedback loop. However, B is not quite sure (lines 20, 24).

In line 25, G asserts that Claim A is the correct possibility and B asks for the evidence.

When G explains that the system is trying to get to a medium (line 28), B asks her if that is the evidence or the reasoning, to which G responds that it is the reasoning. Knowingly or unknowingly, B seems to push G for the evidence for their claim (line 31) and G responds by pointing to the previously collected evidence (line 32).

This reference back to the simulation seems to cause B to be thoroughly convinced as he exclaims, “It’s like feedback!” (line 33). The pair ends their conversation by coming to the consensus that gene regulation is like a home heating system and that it uses negative feedback to accomplish this.

<table>
<thead>
<tr>
<th>Discourse Interaction Excerpt</th>
<th>Commentary</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the second excerpt shown in Figure 4, two pairs of students discuss how enzymes move and they come to consensus using newly collected evidence on the argumentation prompt that used three different analogies of a hungry traveler looking for pizza in a new town.</td>
<td></td>
</tr>
</tbody>
</table>
B1 (boy on left), B2 (boy on right), G1 (girl on left), and G2 (girl on right).

1. B2: ((Reads question aloud.)) How do the enzyme and starch substrate come together to interact? Discuss the following possibilities with your group. Choose the one claim, either A, B, or C, you think is most likely and write down your group’s evidence and the reason for that choice. Run the experiments/simulations as many times as necessary to establish your claim. Claim A: enzymes are drawn to substrates like a hungry traveler without a map in a new town who smells pizza from a distance and heads toward the site.
2. B1: That’s an interesting analogy.
4. G1 and G2: (inaudible, looking at worksheet)
5. B1: Oh God, I find that way too funny…(inaudible)
6. B2: So, claim B? I kinda want to run this again… ((Leans over and clicks several times with the mouse and runs the simulation))
7. G2: I am running it again.
8. B1: ((Looking at the simulation that B2 is currently running)) Huh, intriguing.
9. B2: Well, it’s not like the en … not like the enzymes are following the…ummm…
10. B1: Enzymes?
11. B2: Enzymes… Yeah, they’re not following the starch molecules…they’re just going back and forth.
12. B1: They’re just moving randomly. So, so ok, it can’t be this one… (Pointing to the worksheet) because they are moving randomly…they’re not just following them. Ready, run it for 30 more seconds just to see…ready? ((B1 now takes control of the mouse and clicks on the computer to re-run the simulation…both boys look at the computer screen together.)) Like follow one of these things…they’re not following a particular…or maybe he is!
13. B2: Wait, run that again. ((B1 then clicks to re-run the simulation again.))
14. B1: Follow that one (points to the screen). No, okay, it’s not following. It is completely random.
15. B2: No… so it was following this one (Points to the screen at a specific enzyme) and then it was going and then just completely shot off this way…
16. B1: Yeah…
17. B2: So it is random.

Initially, B2 believes that Claim B (enzymes find substrates by actively looking) is the correct explanation (line 6), but decides to run the simulation again. Similarly, G2 states that she is also running the simulation again.

After having looked at the new evidence together, B1 and B2 decide that the enzymes are just “going back and forth” and “moving randomly” (lines 11-12).

B1 decides that he wants to re-run the simulation and collect more evidence; this time, focusing on following one particular enzyme for more specific information (lines 12-14).

Through this, both B1 and B2 affirm that the enzyme movement is random (lines 15-16). Later on in their discussion, they come to consensus that Claim C is the correct explanation, which is the claim stating that enzymes are like a traveler that wanders the streets in no particular direction, moving randomly until it happens to hit a pizza place. This is the correct claim and a shift from when B2 stated Claim B as the answer in the beginning of the discussion.

**Figure 4.** Discourse interaction excerpt 2 (1:02-3:00) where students are collecting new evidence.

In the next section, we present two cases that show that students’ claim selection strategies may have had an impact on their ability to construct an argument. Table 2 illustrates the strategy used, the claim selected by each student, and their evidence and reasoning rationales. The first group consisting of two students in the discussion excerpt in Figure 4 (B1 and B2) wrote their claim, evidence, and reasoning in response to the prompt, after they
collected new evidence. We can see in the table that they agreed on the claim, which was the correct one. The evidence and reasoning they used to arrive at the claim were also similar. Conversely, the second group wrote their claim, evidence, and reasoning after they worked independently. We can see that their responses to all three CER sections were different. In addition to that, student 1 in the second case did not accurately describe the mechanism by which molecules spread in their evidence and reasoning. We hypothesize that, because students worked independently in Group 2, they did not check and challenge each other’s understanding vis-a-vis the simulation they had previously run.

Table 2: Students’ argumentation responses from unit worksheets

<table>
<thead>
<tr>
<th>Strategy Used to Come to Consensus (Group)</th>
<th>Biology Unit</th>
<th>CER Question Wording</th>
<th>Student</th>
<th>Students’ CER Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect New Evidence (Group 1)</td>
<td>Enzymes</td>
<td>How do the enzyme and starch (substrate) come together to interact?</td>
<td>Student 1 (B1)</td>
<td>Enzymes find substrates, like a traveler without a map, in a new town, wandering the streets in no particular direction, until they bump into a pizza place. Pizza happens to be the only food they like, so they go inside to eat.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Enzymes movement are random</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Student 2 (B2)</td>
<td>Enzymes find substrates, like a traveler without a map, in a new town, wandering the streets in no particular direction, until they bump into a pizza place. Pizza happens to be the only food they like, so they go inside to eat.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The enzymes are randomly moving around the screen, bumping into starch molecules to turn them into sugar.</td>
</tr>
<tr>
<td>Work Independently (Group 2)</td>
<td>Sugar Transport</td>
<td>What is responsible for the spreading out of the molecules you just observed?</td>
<td>Student 1</td>
<td>Diffusion (random/constant motion) is responsible for the random movement of the molecules.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diffusion moves molecules from high concentration to low concentration.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Student 2</td>
<td>Diffusion is responsible for spreading out.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diffusion is random.</td>
</tr>
</tbody>
</table>

Significance of the study

As previously stated, researchers have found that building argumentation skills based on evidence is often difficult for students (Kuhn, 2010), in part due to social dynamics that can influence how students defend their claims using persuasion rather than appealing to the evidence (Berland & Reiser, 2009; 2011). This study fills a gap in the research on scientific argumentation by identifying how students come to consensus on what evidence to use when making claims or constructing explanations. Where we saw a large number of generative/reflective strategies in our study, we hypothesize that this was due to the use of carefully scaffolded inquiry-based computer simulations that required students to engage in sense-making using evidence they collected from experimental iterations. Where we also found a large number of codes in the extraneous/non-reflective categories, we see how a portion of our students were unable to link evidence to their claims, and suggest that teachers need to become aware of these strategies to emphasize and model optimal evidence use.

Furthermore, in response to our question of how reflective and evidence-based strategies can be supported, we offer two possible strategies. First, it may be important to ask students to re-run the simulation while comparing results that were already collected. In the second discussion excerpt, B2 decided to run the simulation again, perhaps in response to following the simple instructions on the worksheet that encouraged students to run the simulation as many times as they wanted. However, as Figure 2 shows, this was not the case for every group. By having a specific task that asks students to re-run the simulation, they can be directed to...
collect new or more evidence to support their claim but this should be emphasized in instruction. Second, similar to B in the first discussion excerpt, assigning an individual in the group as the "evidence manager" or "evidence boss," who directs the conversation of the group to the supporting evidence for the claim they chose, may further remind students to continually engage with the evidence.

Finally, in response to the fourth research question, we identified promising evidence that reflective group consensus processes such as collecting new evidence produced more accurate CER responses compared to non-reflective group consensus processes such as working independently. Fundamentally, science is about explaining the world around us—scientists look at evidence to develop claims and to negotiate their ideas about the validity of each claim (McNeill & Krajcik, 2011). As is true with scientists, if a group of students engage in a “follow the leader” or “work independently” approach, they are not negotiating their ideas and ultimately may not arrive at the best explanation for a particular scientific phenomenon. For this reason, it may be important to name and scaffold reflective group strategies when preparing students to engage in scientific argumentation. However, as these are only two cases of students’ argumentation responses, further analysis of all students’ worksheet responses will need to be conducted in the future, using a topic-specific rubric for each of the units as described by McNeill and Krajcik (2011) in order to see broader, more generalizable trends. Nonetheless, this study has provided some evidence to support the development of more explicit instructional prompts that can direct students to appeal to evidence via scaffolded experiences using dynamic computer supported simulations. We suggest that doing this can support stronger argumentation skills—both in verbal discourse and written explanations.

Endnotes
In the written discourse account, gestures and descriptions of ongoing dynamics are encased in double parenthesis, e.g., (()), direct utterances are written in normal text, and time elapses are marked in single parenthesis with the number of seconds that have gone by, e.g., (5).

References

Acknowledgments
This work was funded by the U.S. National Science Foundation Discovery Research K–12 (DRL 1019228).
Redistributing Epistemic Agency: How Teachers Open Up Space for Meaningful Participation in Science

Mon-Lin Monica Ko, Learning Sciences Research Institute, University of Illinois at Chicago, milko@uic.edu
Christina (Stina) Krist, Curriculum & Instruction, University of Illinois at Urbana-Champaign, ckrist@illinois.edu

Abstract: Taking the practice turn (Ford & Forman, 2006) necessarily involves students in making judgements about the state of their knowledge and in making decisions about how their investigations should proceed. In this study, we investigate how teachers open up aspects of their curricular activities and invite students to partner in the epistemic decisions that drive classroom activity. We draw on work in 3 classrooms as instrumental cases to illustrate what teachers did to ‘open up’ the curriculum and how these moves re-distributed epistemic agency. In each of the three cases, what teachers opened up had different implications in terms of their reach: the resulting “ripple effect” of the impact of the decisions that students were involved making. These cases illustrate the “disciplined improvisational” work (Sawyer, 2004) required to open up space within the constraints and boundaries of the curriculum to create opportunities for re-distributing epistemic agency.

Role of epistemic agency in scientific practice
One of the key challenges of the “practice turn” in science education (Ford & Forman, 2006), captured in part by the Next Generation Science Standards (NGSS Lead States, 2013), is supporting learners’ participation in building explanatory model-based accounts of natural phenomena (Lehrer & Schaufler, 2006; Passmore & Svoboda, 2012). As embodied by this “practice turn,” students should be involved in both making judgements about the state of their knowledge so far and in making decisions about how their investigations should proceed accordingly (Ford, 2015; Stroupe, 2014). Involving students in this way is often cited as a distinguishing factor between “rote” versions of doing science and more meaningful forms of participation in science practices (Barton & Tan, 2010; Berland et al., 2016).

We describe this aspect of this vision for meaningful science learning as supporting students’ epistemic agency. We view epistemic agency as a dynamic and multidimensional construct negotiated through interaction, rather than as a binary property that one either has or does not. More specifically, we view student agency in classrooms as “the way in which [a student] acts, or refrains from acting, and the way in which her or his action contributes to the joint action of the group in which he or she is participating” (Gresalfi, Martin, Hand, & Greeno, 2008, p. 53). We add the modifier of epistemic because we are focused on the actions (or refrains from action) that are consequential to the collaborative construction of a shared knowledge object (Damsa et al., 2010). In our case, this shared knowledge object is an explanatory account of a natural phenomenon. To make the “practice turn,” there needs to be significant redistributions of agency in the classroom, in ways that invite students to partner in the epistemic decisions that drive classroom activity. What this looks like, and how teachers do this in the context of a set curriculum, is the focus of this study.

What makes epistemic agency epistemic?
Actions that we consider consequential to the construction of a shared knowledge object are those that are connected to, or have an impact on, the intellectual nature of the task itself. For example, imagine an engineering challenge designing a fast CO2-powered car. In this challenge, students are charged with designing the car and explaining how and why some cars in this context are faster than others. Students might take actions such as deciding on the type of material used for the body of the car. We would say their doing so indicates agency. But whether this action reflects epistemic agency depends on the rationale behind this decision. For example, a rationale for choosing balsa wood instead of oak because it has less mass and therefore requires less force to speed it up has implications for the explanatory account. We would argue that this action + rationale reflects epistemic agency. However, the same decision with a different rationale could be non-epistemic in nature: if balsa wood is prettier or easier to paint, for example, then the action of choosing balsa wood is not (in that moment) a reflection of epistemic agency.

In the engineering task described above, there are different “levels” of activity that can be opened up for students’ decision-making (Figure 1). For example, when students decide between balsa wood and oak when considering speed, this decision is directly tied to the claims they can make based on that data (Claims, Figure 1).
Students could also be involved in deciding how they should go about investigating that one car is faster than another, such as choosing whether to measure instantaneous speed at a single point, or the average speed across a set or total distance. These decisions influence the methods for investigation (Methods, Figure 1), and the data generated from the trials have a direct impact on the claims that can be generated, supported, or refuted using those data. At yet another “level,” one could imagine students making choices about what the engineering goal is in the first place (Goal, Figure 1), such as deciding whether they are investigating how to make cars fast, or whether they want to make cars that are both fast and use sustainable energy sources. This decision would have implications for the entire duration of the explanatory arc: all the subsequent twists and turns in the models they generate, the methods they use, and the claims they support or refute. Decisions at each of these “levels” invite students into doing substantive intellectual aspects of the work. But they have a different resulting “reach” in terms of the impact on students’ subsequent learning.

Figure 1. “Levels” at which teachers can open up space for epistemic agency within scientific activity.

Teacher decision-making that opens up space for epistemic agency

For students to be able to make any of these decisions, teachers need to do work to renegotiate the positional framing (Shim & Kim, 2018) of students in their classrooms, as these kinds of epistemic decisions are typically considered to be the teacher’s responsibility. Following Hand’s (2012) conceptualization of students “taking up space” in mathematics classrooms, we call this teacher work “opening up space”; creating room for students to be able to make disciplinarily substantive decisions about their science inquiry. In considering when and how to open up space, teachers are fully aware that they themselves have significant responsibilities and constraints that prevent them from giving “full reign” over to students: limited time with students, accountability to content standards, responsibility for student safety and well-being, etc. Therefore, teachers must make strategic decisions about when and how to open up space for students to take on greater epistemic responsibility.

We focus in this paper on the different “levels” at which teachers might strategically decide to open up space for student decision-making, and how opened spaces at those different levels impact students’ subsequent participation in science practices. As such, we focus on curriculum materials as a guiding resource for teachers’ noticing and decision-making about opening up space. We pay particular attention to the consequences of opening up space at each gradient level in terms of both teacher facilitation and students’ subsequent activity.

Curriculum as resource in teacher decision making

Curriculum has historically been designed and used as a resource for shaping science learning in K-12 classrooms. The design principles and features of curriculum materials influence what teachers teach because they specify the
set of activities or phenomena that cohere to a set of learning objectives or student performances. But how lessons are enacted is subject to a great deal of variation. Curriculum enactment is a function of the social interaction between teachers and students, embedded in specific questions or phenomena, and influenced by how schools and districts support and assess teachers and students. Thus, teachers are constantly engaged in the work of adaptation and improvisation (Brown & Edelson, 2003; Sawyer, 2004); how they frame, launch, and enact tasks can directly influence the epistemic goals and knowledge building processes that students engage in (Kang et. al, 2016).

In the context of engaging students in scientific practices, taking the practice turn requires that teachers create room for uncertainty and contention (Ford, 2015; Manz, 2015). It also requires that teachers scaffold students into norms for sensemaking and evaluation. Teachers need to notice and respond to students’ ideas, but also provide opportunities for students play a key role in doing the substantive work of proposing claims and questions and evaluating candidate ideas in the face of emerging evidence (Maskiewicz & Winters, 2012). Although teachers’ responsiveness to students’ ideas and their adaptive use of curriculum materials is necessary, it also creates varying degrees of uncertainty in how science learning unfolds in the classroom. This uncertainty pushes back on the very purpose of clear and tractable learning objectives and outcomes that teachers are accustomed to designing their instruction around. That is, by shifting epistemic agency, teachers subject their classrooms and instructional plans to uncertainty.

In this study, we explore variation how teachers shift epistemic agency by opening up aspects of their classroom activities. The research questions guiding this study are: when and how do teachers create space for students’ epistemic agency? How does that impact students’ participation?

**Methods**

The primary data source for this paper is a corpus of transcripts derived from video recordings of middle school science classrooms collected as part two NSF-funded projects. The goal of both projects was to understand how students learned to participate meaningfully in science knowledge-building practices and how teachers supported that participation over time in the context of the same curriculum, *Investigation and Questioning our World Through Technology*, or IQWST (Krajcik, Reiser, Sutherland, & Fortus, 2011). As a part of that work, both authors conducted classroom observations and generated field notes, analytic memos, and preliminary analyses that focused on how aspects of the enactment differed from other observed classes and from the written curriculum. Multiple lessons (and sometimes, the same lessons) were observed over 3 years.

Guided by our working definition of epistemic agency, we worked to identify candidate moments from our preliminary analyses as instrumental cases (Stake, 1995) in which teachers did significant work to open up aspects of the IQWST curriculum to students’ decision-making. That is, we were looking for moments in which the “opening up” was notable and even surprising to us as observers. We took this approach because our goal is to carefully examine and develop a rich, contextualized description of epistemic agency and the teaching work and tensions involved in supporting it rather than to make claims about, for example, how often or regularly teachers opened up spaces for students’ epistemic agency. We then verified that each of the selected moments were instances in which epistemic agency was re-distributed, from the teacher to the student. To be considered an instrumental case each selected moment had to fit the following criteria:

1. when something that was written as taken-for-granted in the curriculum materials was presented in class as problematic, uncertain, or open to negotiation, OR
2. when teachers added a discussion that was not in the written curriculum materials at all; AND
3. when students responded to the teachers’ invitations in 1 and/or 2 by generating multiple possibilities for claims and provided rationales for those possibilities.

For each of the cases, we then characterized the level at which space was being opened to a redistribution of agency: that is, what was it that there was now room for discussion and deliberation about (see Figure 1)? For each of the 3 identified cases, space was being made around: 1) Phenomena or problem space that are being explored; 2) Methods for exploration, including both whether and how to conduct investigations; OR 3) Claims about the underlying causes of a phenomenon, including both claims drawn from previously constructed models and new observations or conjectures. Taken together, these cases represent variations in what the teacher opened up, and in turn, differences in the levels of epistemic agency for students to take up. For each case, we describe the context of the lesson, what aspects of classroom activity were being opened up to uncertainty, what the teacher did to shift epistemic agency, and how these decisions shifted student participation.

**Findings**

**Case 1: Opening up space for developing methods for investigation**
This Chemistry lesson took place in the Spring semester of the 6th grade IQWST unit centered on the driving question, “How can I smell things from a distance?” The unit spanned multiple weeks and explored the molecular nature of matter. Rather than starting the unit by introducing students to the particle nature of matter, students explore a variety of phenomena to develop a model of how people smell odors. This model is then used to explain the behaviors of solids, liquids, and gases. The model that is generated and revised throughout this unit reflects the understanding that matter is made of particles that are in constant motion.

The focal lesson, as specified in the curriculum materials, asks the teacher to begin with a demonstration that involves two unlabeled flasks of clear liquids. Students are asked to generate observations that can potentially differentiate between the two liquids, and that they “should recognize that further investigation is needed”. The teacher is then asked to demonstrate how an indicator paper, like other detectors explored in lessons prior, detects the presence of different materials by changing colors. The teacher would then carry out a demonstration that shows that the two liquids are different by placing the indicator paper into the two flasks. Because one flask contains ammonia and the other, acetic acid, the indicator paper turns blue and red, respectively. Thus, the indicator provides observable evidence that the flasks contain different liquids.

Instead of following instructions outlined above, Ms. K began the lesson by showing students the two unlabeled containers of clear liquids, and then asked her students, “...how can we tell the difference between these two sources of odor? Who can tell me? What’s different? Let’s put a list on the board here of what’s different between these sources of odor.” After the students identified their similarities (e.g. both liquids were clear) and differences (e.g. one seemed “foamier” than the other), Larry conjectured that the two flasks might contain different liquids. Ms. K then asked how they might test Larry’s claim, saying, “They might be different types of liquids. How can we figure that out? How could we tell?” Rather than jumping in and using the indicator paper to show that the liquids were different, Ms. K’s invitation re-distributes agency by opening up the method of investigating whether the two were different, rather than taking this as a stated fact.

Students took up Ms. K’s invitation by proposing a variety of ideas for investigation, such as smelling, shaking, freezing, weighing, and tasting the liquids (to which Ms. K responded that this was not allowed in the science lab) to determine whether they were the same or different. Another student proposed combining each liquid with another chemical to see how they react, or weighing the two flasks to detect differences in density. Finally, Steven suggested a method of investigation that mirrored how the indicator paper would be used:

Steven: So you could get a turkey baster or a syringe or something, and then put [it] in there, take some of the liquid, and put it on a paper and do the same thing with the other one. And then if something different happens, then you’ll know the difference between what it is.”

It is important to note that Steven not only suggested what materials that could be used (turkey baster, syringe, paper), how they might be used, but also what the results would imply about the liquids. Inviting students into this work has important epistemological implications because students are invited into the work of making decisions about how they might go about investigating the difference between the two liquids, which impacts the type of data that might be used as evidence to support alternative claims.

**Case 2: Opening up space for generating alternative claims and motivating model revision**

This focal lesson comes early in an IQWST unit centered on the driving question “How do my eyes allow me to see?” The unit began with an optical illusion that seeded uncertainty in whether students could really believe their eyes. The beginning lessons explore the behavior and properties of white light as “interacts” with various objects, and progresses to an exploration of how light moves, how light is detected by the eye, and how light is perceived. Over the course of the unit, students co-construct and refine a model of light’s behavior, based on the data from their data collection activities. Notably, this unit is used in the Fall semester, and students’ very first introduction the practices that are emphasized in the curriculum. Thus, there are a number of lessons dedicated to discussing what counts as scientific model, an exploration of how to depict ideas and relationships in model form, as well as the various types of models that count as a scientific model.

In line with the written curriculum, Ms. B began the lesson with a demonstration. She turned off all the lights in her classroom and closed the blinds and aimed a flashlight at a piece of white paper that has been taped to the whiteboard, and asked students to explain why they were able to see the paper. She then directed the beam of light at a handheld mirror, rotating the mirror so that it reflected the light around the room, and, to the students’ dismay, directly at their eyes! In the written curriculum, this initial demonstration was intended to “generate evidence about the difference between the ways light bounces off a mirror and off a [paper]”. The curriculum
specifies that after this demonstration, the class should move directly to gathering empirical data using a light meter, a piece of paper, and a mirror, so that they can use these patterns to describe how light bounces differently off these objects (e.g. scattering, reflecting).

Instead of moving directly to the activity involving light meters, Ms. B engaged her students in a whole-class discussion that spanned half of the class period. During that discussion, she explicitly used the differences in how light behaves with a mirror and a piece of paper to problematize their current model. For instance, when Ms. K directed the beam of light toward the students, who reacted by laughing, grimacing, and talking over one another when the beam of light was aimed directly at them. Ms. B called out this discrepancy in light’s behavior, saying, “When I [aimed the light at the] paper you weren't yelling.” She then asks the students to explain why: “So what makes it different when light bounces…why do you think it’s different from when the light hits your paper and when it hits the mirror?”

Students responded to Ms. B’s invitation by proposing candidate explanations. Adam first proposed that these differences might be happening because light is carrying different information when it bounces off a mirror versus a piece of paper. Later, Illiana weighed in and conjectured that the light is shining off of the paper, but bouncing off the mirror. Although the class agreed that the light was behaving differently, Ms. B pointed out their existing model was insufficient for explaining why:

Illana: when you use paper, it's just shining (inaudible) but when it hits the mirror it’s bouncing…
Ms. B: So why does it do it for a mirror and not for a paper?

(Many hands raised in the front of the room)
Illana: ‘Cuz the mirror is like if we look you can see your reflection...(inaudible)
Ms. B: Why (asking Illana, who seems stumped)...Anne?
Anne: Because that's the mirror and...(inaudible) its cuz the light...(inaudible) light anymore because some light (inaudible)
Ms. B: So is the light hitting different objects differently?
Students in unison: Yes

Ms. B: But how does light know that? How does light know…to act differently when it hits this (pointing the flashlight at the paper) and react differently when it hits this (pointing the flashlight at the mirror)?

Ms. B’s continued press for why encouraged students to propose alternative claims about what might be causing the observed differences in how light behaves with a mirror and paper. These questions emphasized that their current model was unable to account for these different behaviors. Notably, the curriculum materials specify that this demonstration should be used to help students see that light behaves differently, but it does not articulate how these observations problematize the existing model. Opening up their model to uncertainty encouraged students to consider reasons for the different behaviors.

At the conclusion of this opening discussion, Ms. B positioned the data from their upcoming investigation as an opportunity to gather evidence for or against those claims, saying, “Should we get some data, collect some more information about this? Try to figure this out?” Thus, the data generated through their subsequent investigation were positioned as evidence for refining their current consensus model.

Case 3: Opening up space for expanding the explanatory goals
This case took place during the first few days of an 8th grade biology unit organized around the driving question, “Why do organisms look the way they do?” Over the course of the unit, students worked to build a Mendelian model of genetic inheritance. As written, the first lesson in the curriculum had students a) look at a photo of fish and birds and speculate about the functions of their different traits; b) record the frequencies of various human traits amongst their classmates, such as hitchhiker’s thumb and attached vs. detached earlobes; and c) discuss whether these traits were inherited or acquired.

Instead of beginning the unit as written, Mr. M skipped the pictures of birds and fish and had students make observations about the traits of people around them. This motivated a conversation about whether these traits were inherited or acquired, which (unexpectedly) lasted for two class periods. Mr. M had planned for this discussion to then lead into an investigation in which students collected data on a specific trait by observing a specific group of people (e.g., their families; people who came into Burger King over the course of an hour; the
first 20 people whose photos came up on a Google Image search for “celebrities”; etc.). During the discussion about human traits, a debate began about whether height was inherited or acquired when a student made a claim that height was inherited: if your parents are tall, you are likely to be tall, and if they are short, you are likely to be short as well. Several students immediately responded to this claim with counterexamples from their own families: a sibling who was very tall even though their parents were short, or a mother who was short but an aunt who was tall like them. In addition, a few students brought up ideas about nutrition, such as that they should not drink coffee because it would stunt their growth, or that they should drink milk because it would make them grow tall and strong, leading them to argue that it was at least partially acquired. Still other students argued that it was totally random: height was unpredictable based on either your parents or grandparents, or from your nutritional choices.

Mr. M allowed this discussion to continue for nearly the entire 40-minute class period, much longer than in any of his other sections. Towards the end of class, he made a move to transition them into the planned data collection activity (Activity 1.2), still leaving open the possibility of continuing the discussion:

Mr. M: I’m not going to resolve this but I’m feeling like, if this is becoming a compelling question, [...] if we’re going to figure out if something is random, I feel like we should collect some data. So uh. Let’s take a couple more comments and if we’re kind of moving towards that ‘what’s the pattern’ question, then I want to move to [Activity] 1.2. If not, if we’ve just got, if we’re generating some questions, cool, let’s do it now. So um. [points to Carmen] Did you have--?

In response, Carmen brought up how her hair is very similar to her aunt’s but not to either of her parents’ and asked if inheritance could work that way rather than only through parents and grandparents:

Carmen: I actually wanted to make a counterarg—like you said not aunts and uncles, but like I have really wavy and like sometimes curly hair, but like my mom and my sister and my dad all have like the straightest hair, but like my aunt has really curly hair like me, so like, can’t like that be a thing too?

It is important to note here that Carmen’s input was a direct interruption to Mr. M’s bid to “move on” to the next planned activity, and to Mr. M’s insistence earlier in the conversation that students stick to “direct” lines of inheritance (parents and grandparents). It also demonstrates that students were bought into this additional dimension of complexity. In response, Mr. M pointed out it sounded like there was a pattern that a lot of people were noticing: they had traits that didn’t come directly from parents or grandparents. Mr. M then explicitly acknowledged the push-back they had been giving him and allowed space for the four students who still had their hands up to share their ideas:

Mr. M: Let’s here, so, I’ve been trying to pull us in a certain direction, I like where you guys are pulling us though, I like that you’re wanting to stick with examples from our own families. In our last four I want us to make sure that you’re turning to our speakers [...], so I want us to face this way and let’s go to Elena.

The bell rang after Elena shared. Before they left, Mr. M gave each student an index card and asked them to record the patterns from their families that they thought were important “clues” to helping them understand inheritance. When students returned the next day with their index cards with family patterns, they made a second driving question board with those “family cases.” As they were constructing this driving question board, the discussion about whether each trait mentioned was inherited or acquired continued, including conversations about traits such as athleticism, spoken language, and race. After another entire class period of this discussion, Mr. M asked students to pick one of these traits to observe and to pick a population in which to observe it over the course of the weekend (Activity 1.2 as planned).

Mr. M’s decision to allow the classroom conversational space to be opened to students’ persistent interest in the patterns in their own families, and his in-the-moment response to create a second driving question board, broadened and made more complex the set of phenomena that students were trying to explain. In addition, his doing so allowed students to jointly decide, through conversation, what kind of explanation they were going to be working to build over the course of the unit. This was no longer a unit about different traits in different species, or quirky differences between humans like hitchhikers’ thumbs. Instead, it was a unit about complex patterns of inheritance within their own families—the thing that was both interesting and bothersome to them. It set the expectation that whatever models they were developing should be able to at least partially explain these complex
patterns, and that whatever their models could not explain should drive their future investigations. And that was the case: students continued to draw on family-based examples throughout the unit, even while discussing the genetics of plants and moths.

Discussion

In each of the three cases, what teachers opened up had different implications in terms of their reach: the resulting “ripple effect” of the impact of the decisions that students were involved making (Table 1). Ms. K opened up the methods for investigation in Case 1, creating the opportunity for students to shape how they conducted the investigation and the claims they could draw from it. In Case 2, Ms. B problematized their existing model of light, which opened up space for students to make alternative claims. These decisions framed their subsequent investigations for the next few days, leading to revisions to an aspect of their light model. Finally, in Case 3, Mr. M opened up the opportunity to expand the phenomenon that anchored the entire unit. The “reach” of this decision was long-ranging: even though his students completed most of the activities in the unit as written, they were framed by, and constantly connected to, a need to explain the puzzling patterns of inheritance that students observed in their families.

Table 1: Overview of cases

<table>
<thead>
<tr>
<th>Case 1 (Ms. K)</th>
<th>What aspect is the teacher opening up?</th>
<th>How do students participate?</th>
<th>Reach</th>
</tr>
</thead>
<tbody>
<tr>
<td>The methods for investigation</td>
<td>Students brainstorm candidate methods &amp; tools that will generate evidence to help distinguish two unknown liquids</td>
<td>Within the investigation in a single lesson</td>
<td></td>
</tr>
<tr>
<td>Case 2 (Ms. B)</td>
<td>Their current explanatory model for light</td>
<td>Students propose alternative claims for the observed phenomenon. This leads to a new phase of evidence collection and subsequent model revision in later lessons</td>
<td>Investigations and model revision within days of a single unit</td>
</tr>
<tr>
<td>Case 3 (Mr. M)</td>
<td>Anchoring phenomenon that motivates explanatory goal</td>
<td>Students continued to bring up puzzling patterns of trait inheritance in their own families, which eventually become part of the anchoring phenomenon.</td>
<td>Influences explanatory goals (and models, investigations, and claims) for the entire unit</td>
</tr>
</tbody>
</table>

Importantly, none of these cases were instances where students did the “opening up” spontaneously or intentionally (cf. Calabrese Barton et al., 2013; Engle & Conant, 2002). This was purposeful on our part. When teachers open up space for students to make choices about acting or not, and how that action contributes to the joint action of a group (Gresalfi et al., 2008), they also subject their own instructional practice to uncertainty—about what students will bring to the table, whether it fits in with their instructional goals, etc. These cases describe the “disciplined improvisational” work of teaching (Sawyer, 2004). They illustrate how teachers can open up space while still maintaining some of the constraints and boundaries of the curriculum (and therefore limiting the degree of uncertainty they are introducing to their teaching) to create opportunities for re-distributing epistemic agency. Seeing aspects of their curriculum—the methods, models, and anchoring phenomena—as entry points for re-distributing epistemic agency may help teachers make inroads to shifting their classroom instruction towards more responsive instruction: not all teachers may feel capable of shifting their instructional practice to look like Mr. M’s at the start, as it requires substantive changes to in one’s instructional sequence. But they may be able to take a step towards it on the level of a single lesson, like Ms. K. The language of “opening up space” and considering the resulting “ripple effect” may scaffold productive ways of talking about these goals with teachers as we work to increase opportunities to re-distribute epistemic agency in science classrooms.

Highlighting this as an area for supporting teacher learning suggests that it may be productive to think about how we might develop educative curriculum materials that support these teacher learning goals. In addition to getting teachers to attend to places in their curriculum that they can invite students’ participation in ways that helps students to develop accurate content knowledge, we should also be considering whether and how to design materials that focus teachers’ attention on ways to increase students’ ownership and agentic participation in knowledge-building work. For example, as in the case of Ms. B, this could be positioning data collection activities as opportunities for students to apply their knowledge to design procedures and predict their outcomes, rather than an exercise in obtaining a target outcomes. Repeated attempts at creating space for epistemic agency may scaffold teachers toward accepting uncertainty as productive tension for teaching.

The cases presented here are exemplary cases of how teachers hold the need to meet specific learning objectives in tension with the desire to provide students with opportunities to make meaningful and consequential
decisions, and how they navigate both goals well. The decision-making involved in navigating this tension is important to address and support in professional development and teacher preparation. Generating language around these improvisations can be fruitful beginnings for thinking about how we might support teachers to increase opportunities to re-distribute epistemic agency.

**References**


Brown, M., & Edelson, D. (2003). Teaching as design: Can we better understand the ways in which teachers use materials so we can better design materials to support their changes in practice. LETUS Design Brief). Evanston, IL: The Center for Learning Technologies in Urban Public Schools.


**Acknowledgments**

This research was funded by the National Science Foundation Grants ESI-0227557, ESI-0439493, ESI-0101780, and DRL-1020316, and by the NAEd/Spencer Dissertation Fellowship Program. Any opinions, findings, conclusions, or recommendations expressed here are those of the author and do not reflect the official views of the organizations above.
Examining the Role of Explicit Epistemic Reflection in Promoting Students’ Learning From Digital Text
Feng Lin, Dana Gnesdilow, Nicole D. Martin, Catherine L. Dornfeld Tissenbaum, and Sadhana Puntambekar
feng.lin@wisc.edu, gnesdilow@wisc.edu, ndmartin@wisc.edu, cldornfeld@wisc.edu,
puntambekar@education.wisc.edu
University of Wisconsin-Madison

Abstract: This study examined whether and how epistemic reflection embedded in students’ use of a digital tool influenced their learning using the tool. One biology teacher and his four classes of 8th graders (N = 100) participated in this study. Students used a digital text tool, VidyaMap, to learn about photosynthesis and energy transformation. Two classes were provided prompts for epistemic reflection. The prompts encouraged students to reflect on the epistemic role of the tool for inquiry. The other two classes served as comparison groups and used the tool without epistemic reflection prompts. Quantitative analysis showed that the classes that received the epistemic reflection prompts outperformed the comparison classes in their learning. We also coded the levels of students’ epistemic reflection and found that it was correlated with students’ science learning from the tool. Qualitative analyses further suggested how students with high and low epistemic reflection scores differed in their inquiry using the tool. Implications of findings are discussed.

Introduction
Since Perry’s work (1970), epistemic cognition, understanding of the nature of knowledge and knowing, has attracted much attention in research. Epistemology examines the origin, nature, methods, and justification of human knowledge (Hofer & Bendixen, 2012; Hofer & Pintrich, 2002). Over the years, researchers have adopted a psychological approach to examine individual’s epistemology, focusing on what individuals believe is the nature of knowledge and knowing (Hofer & Pintrich, 1997; Khine, 2008). Accumulated evidence has shown the role of epistemic cognition in students’ learning processes and outcomes (Cano, 2005; Mason, 2010; Mason, Ariasi, & Boldrin, 2011; Qian & Alvermann, 1995; Schommer, 1990). However, much of the research was based on self-reported questionnaire for examining students’ epistemic understanding, which has been criticized for its decontextualized nature (Mason et al., 2011) and limited explanatory power (Sandoval, 2012). More recent literature has called for studying epistemic cognition from a situative perspective (Chinn, Buckland, & Samarapungavan, 2011; Louca, Elby, Hammer, & Kagey, 2004; Sandoval, 2012). The situated view argues that epistemic cognitions are situated and context-dependent (Chinn et al., 2011; Louca et al., 2004), and that they “emerge from and are linked to particular forms of activity” (Sandoval, 2012). Given the situated nature of epistemic cognition, it is possible that students possess certain epistemic ideas, but they are not activated in certain contexts (Louca et al., 2004), which may influence how they learn. Therefore, an important question to ask is, how can we design epistemic support to activate students’ epistemic cognition, and will this support help students learn better? This study will address this issue in a context where digital tools are used.

With the widespread use of technology in education, digital tools are commonly used for supporting students’ inquiry and learning. Many of these tools have been designed with epistemological underpinnings that may not be obvious to or taken up by learners. Therefore, students may not use the tool as intended, which might influence their learning. In this study, our aim is to examine whether engaging students in epistemic reflection, that is, providing students with opportunities to reflect on the epistemic role of tools used to support science inquiry, could improve their learning. We adopted the term epistemic reflection from Mason et al. (2011), who examined students’ spontaneous thinking about knowledge and knowing. But here, we use it to refer to the intervention process of activating students’ thinking about knowledge and knowing. In this case it is to explicitly engage students in reflecting on the epistemic role of a digital tool for their inquiry. Engaging students in such a process might influence how they actually engage in inquiry in that context, and therefore influence their learning. Our work builds on previous studies that emphasized the importance of making the epistemic aspects of students’ inquiry explicit (Lin & Chan, 2018; Sandoval & Reiser, 2004; Schauble, Glaser, Duschl, Schulze, & John, 1995). For example, Sandoval and Reiser (2004) designed epistemic scaffolds to structure students’ inquiry activities within ExplanationConstructor (an electronic journal that students used to record their investigations). Their aim was to help students attend to the epistemic features of scientific explanations, including articulation of coherent, causal accounts, and use of data to support causal claims. They found that such an epistemic tool was helpful for improving students’ inquiry. Schauble and colleagues (1995)
found that 5th grade students designed better experiments after instruction on the purpose of experimentation. These studies suggested the importance of making epistemic goals and epistemic aspects of students’ inquiry explicit. The current study builds on this line of research. To help students attend to the epistemic features of the tool, we used explicit epistemic reflection prompts in an attempt to activate students’ cognition about the epistemic role of a digital tool for inquiry and examined its effect on students’ learning using the tool.

The particular digital tool we focused on in this study is VidyaMap. VidyaMap uses both concept maps and text to facilitate students’ navigation and inquiry (Puntambekar & Stylianou, 2005)[see Figure 1]. The concept maps display the connections between science ideas, which mirrors the interrelated structure of the science concepts and phenomenon. Underlying this design is the epistemic idea that scientific knowledge is coherent and connected. This epistemic feature of VidyaMap was designed to help students engage in deeper and more sustained inquiry, as the connection of concepts could be used to guide and expand students’ inquiry by helping them generate more questions and leading them from one idea to another. However, students may not understand this epistemic feature of the tool and the underlying design principles, and therefore may not use it as intended. Therefore, we developed epistemic reflection prompts to activate students’ epistemic awareness of this feature to see if it would affect their learning from their inquiry using the tool. To examine the effects of the epistemic prompts on students’ learning, we also included a comparison group who were not given prompts for epistemic reflection while using the tool.

In addition to exploring the effects of epistemic reflection on students’ learning from using VidyaMap, we examined whether and how the levels of students’ epistemic reflection may be related to the depth of their learning. Previous studies have suggested that students’ epistemic understanding is related to their conceptual understanding (Qian & Alvermann, 1995; Schommer, 1990), but many of them were based on a decontextualized questionnaire approach for examining epistemic cognition. This study will take a situated perspective to examine epistemic cognition by looking at students’ reflections of the epistemic role of the digital tool for inquiry, and examine the relationship between epistemic reflections and student’s learning from using the digital tool. We will also explore how students with different levels of epistemic reflection engage in inquiry using the digital text tool.

Three research questions are addressed in this study:

1) Does the process of engaging students in epistemic reflection improve students’ learning when using the digital text tool?
2) What is the relationship between students’ level of epistemic reflection and their learning?
3) How do students with different levels of epistemic reflection engage in inquiry using the digital tool?

Figure 1. An example of a navigation page on VidyaMap.

**Methods**

**Participants and context**

One experienced biology teacher and his four classes of 8th grade students participated in this study. Each of the four classes was randomly assigned to one of the two conditions: epistemic reflection condition and non-epistemic reflection condition. Students in two classes (n=49) were provided with epistemic prompts (epistemic reflection condition), and the other two classes served as comparison classes (n=51) (non-epistemic reflection condition). Despite the fact that the classes were randomly assigned to one of the conditions, each student was
not randomly assigned to a particular condition, which made this a quasi-experimental design. This study occurred during students' regularly scheduled science classes in a Mid-sized Midwestern city in the U.S.A. The study was conducted within a larger research project that examined how engaging students in a curriculum focused on solving 21st century bio-engineering problems would help them learn science. In one of the units, Make Your Own Compost!, students needed to use information they learned about energy transformation, matter cycling, ecosystems, decomposition and human impacts on the environment to solve the challenge. Thus, all the participants were involved in solving a design challenge to reduce the amount of waste going into landfills by designing compost that breaks down quickly and contains lots of nutrients. Students worked in groups of three to four students throughout the unit and participated in cycles of inquiry which included research using simulation and physical experiments, as well as second-hand research (Palincsar & Magnusson, 2001) using a digital text tool, VidyaMap. Throughout the unit, every student wrote notes in his or her scientist’s journal, which was a paper-and-pencil tool that included prompts associated with each activity, such as writing hypotheses, collecting data, interpreting results, and reasoning. The journal both served both as a scaffold, as well as a place for students to record their ideas.

Procedure
About halfway through the unit, students were engaged in a mini-unit to help them understand the role that plants play in the transformation of energy in ecosystems. Students were asked to: 1) brainstorm their ideas about what plants need to grow, which they recorded in their journals and served as prior ideas for our analysis; 2) write questions they had about plants' role in energy transformation in ecosystems; and 3) use VidyaMap to research their questions. Students' notes from their research on VidyaMap were recorded in their student journals and served as their post ideas for our analysis. It should be noted that all students had used VidyaMap before this session, and thus they were familiar with the concept map structure of the interface.

The prompts were given after students’ brainstorming of ideas and before their generation of questions and use of VidyaMap. For the epistemic reflection condition, each student was provided with a worksheet containing three questions: 1) You already have some experience with VidyaMap, what do you think about the role of the maps (left side of the screen) in VidyaMap? 2) Did the maps help you? What did they help you with? 3) How can the connections between nodes in the maps help you with your research? They were first asked to reflect and write their ideas about these three questions on the worksheet. Then they were asked to discuss these questions in small groups for 10-15 minutes. Students were able to revise their responses to the questions based on their discussion. A similar procedure was followed in the comparison classes (non-epistemic reflection condition), except that we gave them a general prompt: What do you want to learn from VidyaMap?

Measures
Students’ learning from text
To examine students’ learning from text, we analyzed students’ pre- and post- ideas in their journals. Students wrote what they knew about plants in their journals before their use of VidyaMap, these ideas were analyzed as pre-ideas. They also wrote what they learned after their investigation with VidyaMap in their journals. These responses were analyzed as post ideas. We analyzed these pre- and post-ideas with a coding scheme focusing on the extent to which the students discussed the concepts of photosynthesis and energy transformation. The concepts students learned from engaging in these photosynthesis activities were an essential part of the Compost unit, because students needed to understand the role that plants play in transforming energy and cycling matter in an ecosystem to solve the challenge. The coding scheme was inductively developed. For this rubric, the concept of photosynthesis included 6 sub-ideas and energy transformation included 2 sub-ideas, as Table 1 shows. We counted the number of sub-ideas included in each student’s pre- and post-written responses according to the rubric. The total number of sub-ideas included in each response was summed to represent the thoroughness of their ideas about photosynthesis and energy transformation. Thirty percent of the students' journal entries from their VidyaMap session were individually coded by the first and second authors. Cohen’s Kappa was $K = .77$, indicating good agreement. The first author coded the remaining responses.

Table 1: Coding scheme for pre- and post-ideas

<table>
<thead>
<tr>
<th>Concept</th>
<th>Sub-Ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosynthesis</td>
<td>included sun/sunlight, water, and oxygen as input for photosynthesis</td>
</tr>
<tr>
<td>(max 6 points)</td>
<td>included oxygen as output for photosynthesis</td>
</tr>
<tr>
<td></td>
<td>included glucose/chemical energy/food/carbohydrates as output for photosynthesis</td>
</tr>
</tbody>
</table>
explain the role of stomata in photosynthesis
explained the role of chlorophyll/chloroplast in photosynthesis
explained the role of root/leaf/stem in photosynthesis

<table>
<thead>
<tr>
<th>Energy Transformation (max 2 points)</th>
<th>connected plants with sun/sunlight for energy transformation, or state that plants play an important role in energy transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>connected plants with the rest of the ecosystem (e.g., provide food for consumers)</td>
</tr>
</tbody>
</table>

*The maximum score that could be earned was 8.

Students’ reflection on the epistemic role of VidyaMap
To examine students’ epistemic cognition embedded in their epistemic reflection, we analyzed students’ written responses on their prompts worksheet, which served as both an intervention and a measurement. We coded these responses according to their levels of understanding about the epistemic role of VidyaMap. For our purpose in this study, we described a more sophisticated epistemic reflection as one that showed a better understanding of the epistemic role of VidyaMap for sustained inquiry, e.g., how the connections between the nodes in the concept maps might help them expand their ideas for further inquiry. A less sophisticated reflection showed a more superficial understanding about the function of the tool, e.g., it provides content knowledge. Each student’s responses to the three questions were grouped together and coded. Three levels of epistemic reflection were identified as follows in Table 2 below. Thirty percent of the students’ epistemic reflection responses were individually coded by the first and second authors. Cohen’s Kappa was $K = .69$, indicating good agreement. The first author coded the remaining responses.

<table>
<thead>
<tr>
<th>Level of Response</th>
<th>Description</th>
<th>Student Example*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Unintelligible responses</td>
<td>1) It’s the food or (mabbs) to more of the plant parts. 2) Yes it showed me what to do. 3) If you click it it tells you?</td>
</tr>
<tr>
<td>1</td>
<td>Described VidyaMap as tool to find content knowledge and details</td>
<td>1) I think they are really helpful when it comes to the research part of the project. It gives us the background information. 2) It helped me with finding specific details. 3) It gives us the definition of the role all of the objects.</td>
</tr>
<tr>
<td>2</td>
<td>Described connections in VidyaMap as important, but did not discuss the role of connection for inquiry/understanding</td>
<td>1) The role is to show how the maps are all connected and they all spread. 2) Yes helped me see that they are all connected to one [main] thing producers. 3) to show that they are all connected to one thing.</td>
</tr>
<tr>
<td>3</td>
<td>Described importance of connections in VidyaMap for inquiry to include expanding ideas, leading one subject to another, spurring more questions, which were beyond finding information/answers</td>
<td>1) It helps you find the [origin] of the node. 2) Yes, It helps us get more questions and answers. 3) It lets us know [there is] more parts of it instead of just one answer. .... It helps you realize that there are connections between them all. Each thing leads up to another + back sometimes to one main thing. It really helped me realize info[information], that I wasn’t yet aware of.</td>
</tr>
</tbody>
</table>

*We bolded words in students’ examples to highlight key parts of response in relation to the rubric.

Small group inquiry during navigation
Students’ small group discussions during their investigation on VidyaMap were qualitatively analyzed to understand how students with different levels of epistemic reflection may engage in inquiry differently using the tool. We selected two contrasting groups based on their epistemic reflection scores (one high and one low epistemic reflection group) as well as availability of video data. The groups’ epistemic reflection scores were calculated by averaging the epistemic reflection scores of all students in the group. Contrasting cases analysis (Rummel & Hmelo-Silver, 2008) was conducted to examine their inquiry practice using the tool.

Results
Comparing students’ learning gains between conditions
We first conducted a t-test to compare students’ prior biology knowledge in both conditions (epistemic reflection vs non-epistemic reflection condition) using a content test to examine if there were significant differences between the groups. All students took the Compost content knowledge test before the start of the Compost unit, which measured what they knew about energy transformation, matter cycling, decomposition, ecosystems, and human impacts on the planet. The test included 22 multiple choice questions and 4 open-ended
questions. We found that students within each condition had similar levels of biology content knowledge at the start of the unit ($t = -1.197, p = .234$). By establishing that students in both conditions started out with similar levels of content knowledge, we could then assume that the results of our analyses were likely not due to differences in students’ prior knowledge.

We then conducted an ANCOVA to examine if the students in the epistemic reflection condition had higher post-idea scores than the students in the non-epistemic reflection condition after controlling for their prior idea scores. The assumption of homogeneity was met. The result showed that the students in the epistemic reflection condition (adjusted mean $= 3.26$ ($SE = .282$)) had significantly higher post-idea scores than the students in the non-epistemic reflection condition group (adjusted mean $= 2.36$ ($SE = .279$)), $F(1, 94) = 4.98, p < .05$ when controlling prior idea scores. This suggested that students from the epistemic reflection condition learned more from VidyaMap than students from comparison condition.

The relationship between epistemic reflection and students’ learning outcomes

To understand whether students’ epistemic understanding of the online tool in their reflections was related to their learning, we conducted a correlation analysis to examine the relationship between students’ epistemic reflection scores and their post-idea learning scores. The results showed that they were significantly positively correlated ($r = .316, p < .05$), suggesting that more sophisticated epistemic reflection was associated with a better learning outcome.

Students’ inquiry during navigation

To understand how epistemic reflection was related to students’ inquiry and learning, we compared the discourse of two contrasting groups from the epistemic reflection condition: high and low epistemic reflection group. The average epistemic reflection scores of all groups ranged from 1 to 3. We chose the highest ($M = 3$) and the second lowest one ($M = 1.5$) as the contrasting groups as the video data of the lowest one was not available. The Critical Incident technique (Flanagan, 1954) was used to identify excerpts from the groups’ discourse that best exemplified the different types of epistemic practice of the contrasting groups as they conducted inquiries on VidyaMap. Overall, we found that students’ in the high and low epistemic groups engaged in qualitatively different inquiries, which aligned with their respective epistemic reflections. In the following, we present excerpts from both groups that illustrated how they differed in their inquiry practice. The following is an episode from the high epistemic reflection group:

\[
\begin{align*}
S2: & \text{ Like yesterday… We found energy transformation taking sunlight and converting it into chemical energy. So-} \\
S1: & \text{ How do they transform energy? How do they provide energy for the other animals?} \\
S2: & \text{ So we should go to photosynthesis right?} \\
\ldots \\
S2: & \text{ [reading VidyaMap page] It says, uhm, the- there’s sunlight needed to combine with the nutrients to produce sugars for the plants to grow.} \\
S3: & \text{ Wait!} \\
S1: & \text{ That’s growth, that’s not energy.} \\
S3: & \text{ [pointing to the VidyaMap page] What is that? What is that? Chloroplast- plast.} \\
S2: & \text{ Click on it.} \\
S1: & \text{ I did.} \\
S3: & \text{ Of the plant cells and the-} \\
S2: & \text{ Oh it makes cell membranes like, right there there’s one in the map.} \\
S3: & \text{ Ok. So, this thingy is- oh yeah that’s the one I was talking about! The one that turns the plants green.} \\
\ldots \\
S2: & \text{ Mhmm. But we know- we have to know… why do they transform, energy? So we gotta go back to producers.}
\end{align*}
\]

This excerpt shows how this high epistemic reflection group decided their navigation direction based on their prior understandings and questions they wanted to know, and how they made use of the epistemic feature of the tool (connection of concepts) to expand their ideas. As the excerpt shows, while S2 was reading
aloud about the process of photosynthesis on VidyaMap, S3 noted a related concept (chloroplast) from the visualization of the concept map on VidyaMap, which prompted the group to learn more about chloroplasts.

Later the group was wondering how and why plants make sugar. While S1 was navigating, S2 noted some related information on VidyaMap:

S2: [pointing to VidyaMap page] Right there. It says um photosynthesis is the ability to convert sunlight into carbon and sugars- ((continues reading text from VidyaMap))
S4: (inaudible)
S2: For energy.
S1: ((reading quietly))
S2: Go to carbohydrates
...
S2: [talking to S3] What do you think?
S3: Well I just read that um energy cannot be created or destroyed so plants can’t create energy instead they (transform) the energy.
S2: From the sun light, chemical energy. Perfect. ….. ((writing notes and thinking aloud)) Plants can’t create energy but it is transformed-

This excerpt showed how this high epistemic reflection group navigated to different related concepts guided by their emerging questions, and how they connected each other’s ideas (S2 built upon S3’s ideas), and came to understand the relationship between plants and energy transformation.

In contrast, the low epistemic reflection students did not use any questions to guide their inquiry, even though they were asked to do so. They often simply opened a VidyaMap page, one student read aloud, and the rest of the group followed and recorded the information in their journals. For example, in the beginning of a session, the group opened a page on producers:

S1: [reading a VidyaMap page] Plants use the nitrogen and the carbon, well the plants, take, in, carbon and water, and nitrogen that is found in the water, they break down the nitrogen so that other plants can use it.
S2: So what are we writing down?
S1: Take, taking –
S2: Carbon –
S1: Water. Water slash…
S2: Carbon then what?
S1: Carbon, and turn the, turn the nitrogen into usable nitrogen. Carbons into, turn the nitrogen atoms…into usable nitrogen…for other animals. And turn, carbon...into, into food and oxygen.

After a while, another student learned something from another VidyaMap page, and read aloud “Plants are able to make their own food, cause that’s, that’s what they do.” He jotted down these notes, and everyone else in the group also recorded this information in their journal. At times, one student even explicitly told everyone to copy the information he found:

S3: [pointing to a graph in a VidyaMap page] So does everyone write down these four things?
S4: Yes.
S3: Have you?
S1: Two of them.
S3: Write it down.
S4: Come on dude.
S1: Okay (mumbling)).
...
S3: [Reading a VidyaMap page] So six carbon dioxides and six water, plus sunlight...is that equal? One carbon, one carbohydrate and six oxygen. Yeah it does. So write everyone, write this down. Write that equation down. Six, so write underneath your papers...
S4: That means you too XX,
As these examples show, the high epistemic reflection group used VidyaMap as an epistemic tool to support and expand their question-driven inquiry, and they actively connected their previous ideas as well as each other’s ideas for constructing new understanding. However, the low epistemic reflection group merely used VidyaMap as a tool to provide information and did not capitalize on the epistemic features of the tool to help them connect and expand their ideas. Their navigation on VidyaMap was fragmented. They neither connected their search with any research questions nor to their previous ideas. They just copied what was written on VidyaMap and did not process nor connect each other’s ideas. Such differences in inquiry practice between these two groups were aligned with the difference in their epistemic reflection.

Discussion
In this study, we examined the role of embedded epistemic reflection in students’ learning from digital text. Drawing from the previous research emphasizing the situated nature of epistemic cognition (Chinn et al., 2011; Sandoval, 2012), we designed explicit epistemic reflection prompts to activate students’ cognition about the epistemic role of an online tool for inquiry and examined its impact on students’ learning and inquiry using the tool. We specifically addressed three questions: 1) Does the process of engaging students in epistemic reflection improve students’ learning when using the digital text tool? 2) Is there a relationship between students’ levels of epistemic reflection and their learning from using the tool? and 3) How do students with different levels of epistemic reflection engage in inquiry using this tool?

In traditional classrooms, science has generally been presented as a collection of unrelated facts and ideas. VidyaMap was intentionally designed to display the connections between science concepts and ideas in concept maps. These maps can be used by students to help them to see how science concepts are related and to support their navigation in the tool for sustained inquiry. However, this epistemic feature is not always obvious to students. Our study showed that having students explicitly reflect on the epistemic role of the tool better promoted their learning when using it.

Research over the past few decades has been focused on understanding the nature (e.g., dimensionality) of epistemic cognition and its relation to other constructs. Few studies have examined how epistemic support could be designed in a certain context, and how it may impact students’ learning. Our study was situated in a collaborative learning environment where a digital tool was used. Our findings suggest that helping students understand the epistemic nature of the tool by providing reflection prompts may be a promising way to promote students’ inquiry and learning. This finding aligns with previous studies (Sandoval & Reiser, 2004) that emphasized the importance of scaffolding the epistemic aspects of students’ inquiry. It also sheds light on a new way to provide explicit epistemic support in a context where a digital tool is used to impact students’ learning.

To further understand the relationship between students’ understanding of the epistemic role of the tool and their learning, we conducted a correlation analysis to examine the relationship between the students’ epistemic reflection scores and their post-science ideas. Consistent with the previous research (Mason et al., 2011; Qian & Alvermann, 1995; Stathopoulou & Vosniadou, 2007), we found that these two variables were positively correlated. This result confirmed the importance of helping students engage in epistemic reflection to support their learning.

In examining how students’ epistemic reflection might influence their inquiry and learning, we focused on how two contrasting groups with high- and low-level epistemic reflections pursued their inquiries using VidyaMap. We found that the group with a higher-level epistemic reflection attended more to the epistemic role of the tool, and used the connection of nodes to expand their inquiry. They also connected what they knew, what they wanted to know, as well as to each other’s ideas during their discussion. In contrast, the group with lower-level of epistemic reflection did not attend to the epistemic feature of the tool, and simply used VidyaMap as a tool to provide content knowledge. This result showed that students’ understanding of the epistemic features of the tool aligned with their use of it. This is consistent with the previous studies on the relationship between epistemic cognition and learning process (Cano, 2005; Cho, Woodward, & Li, 2017). Many of these previous studies are based on questionnaire measures, our study provided further evidence by examining epistemic cognition from a situated perspective and illustrated how students’ epistemic reflection might be related to their inquiry and learning while they used VidyaMap. It is possible that our epistemic prompts activated students’ cognition about the epistemic role of the tool, which influenced their inquiry during navigation, and therefore their learning using the tool. Future research could further examine this relationship and test it in a different context.

References


Acknowledgments

We thank the students and the teachers who participated in this study. The research reported in this study is supported by a grant from the National Science Foundation to the fifth author (Award Number: 1418044).
Bringing Practices of Co-Design and Making to Basic Education

Sini Riikonen, University of Helsinki, sini.riikonen@helsinki.fi
Pirita Seitamaa-Hakkarainen, University of Helsinki, pirita.seitamaa-hakkarainen@helsinki.fi
Kai Hakkarainen, University of Helsinki, kai.hakkarainen@helsinki.fi

Abstract: The purpose of this study was to analyze five student teams’ (Grade 7) co-design processes that involved using traditional and digital fabrication technologies for inventing, designing, and making complex artifacts. A methodological framework for analyzing maker-centered learning, by relying on ethnographic video data and participant observations, was developed. The study examined the extent to which young students are able to productively participate in creative design and making activities. The results indicated that four of the five student teams successfully engaged in the co-invention processes. The importance of a shared epistemic object of co-design was prominent on every team. Some teams experienced challenges in organizing collaborative processes and the team size appeared to have a significant effect in this regard. The successful teams were able to take on complex and multifaceted epistemic and fabrication-related challenges.

Keywords: co-design, knowledge-creation, maker culture, epistemic objects, teamwork, co-invention

Introduction

The purpose of the present study was to analyze five Grade 7 student teams’ co-design processes that involved using traditional and digital fabrication technologies for inventing and making complex artifacts. The study involved developing a methodological framework for analyzing maker-centered learning as well as examining the nature of the student teams’ co-design processes. Productive participation in the emerging innovation-driven knowledge-creation society appears to require cultivation of sophisticated innovation competencies by all citizens, starting from an early age. Concurrent educational practices do not, however, sufficiently foster young peoples’ creative competencies, because of the strong focus on the transmission of pre-given content knowledge and routine procedures to students. To overcome such limitations and make school learning a more inspiring experience, our Laboratory of Co-Inquiry, Co-Design, Co-Teaching and Co-Regulation (Co4-Lab) project aims to bring elements of the maker culture (see Anderson, 2012) to Finnish schools. To that end, students are engaged in collaborative efforts to invent complex artifacts, sparking intellectual, technical and aesthetic challenges. In this paper, the term “co-invention” is used to characterize artifacts created in students’ knowledge-creation projects, consisting of intertwined co-design and making processes. Exceptional Finnish craft education, technology education, and science-lab infrastructure provide a solid foundation for integrating such maker practices with core curricular activity. We are creating high-end makerspaces in Finnish schools by expanding crafts classrooms with instruments of digital fabrication, such as 3D CAD, robotics, electronic circuits, and wearable computing (e-textiles), with which one may create multi-faceted complex artifacts (cf. Blikstein, 2013). This study examined the co-design processes of small teams of students aged 13 to 14, who participated in the Co4-Lab co-invention project at Aurinkolahti basic school, located in Helsinki, Finland. In this article, we analyze and describe to what extent and how the student teams participated in the co-design processes. In the following paragraphs, we will first present the theoretical framework of our investigations. Then we will go through the research setting and methods of data collection and analysis. Finally, we will present our results and discuss the significance of the findings.

Creating knowledge through artifact-mediated co-design processes

The present investigation relies on our longstanding effort to cultivate knowledge-creating learning that, beyond knowledge acquisition and social participation, involves systematic collaborative efforts in creating and advancing shared epistemic objects by means of externalizing ideas and constructing various types of intangible and tangible artifacts. Learning by making entails, in accordance with Papert’s (1980) constructionism, that learners are not only building conceptual knowledge but also using digital instruments to design and make materially embodied artifacts and cultivating new ways of thinking and acting during the process. Maker-centered learning involves students externalizing their ideas through conceptual (spoken or written ideas), visual (drawing, graphs) or material (3D prototypes and models) artifacts, creating an opportunity for themselves and their peers to build on these ideas, discuss and elaborate on them and embody ideas in progressively more advanced artifacts. Using sophisticated digital fabrication instruments for creating artifacts may be interpreted as providing material agency
for pursuing more complex and challenging inquiries than the participants would otherwise be able to accomplish. We consider knowledge creation as a practical communal activity that, to a significant extent, relies on operational methods, creative processes, and practices (“knowledge practices”) that a learning community can appropriate and cultivate with adequate facilitation, guidance, and real-time support. Artifact-mediated knowledge creation is an emergent and nonlinear process whereby the actual goals, objects, stages, digital instruments or end results cannot be pre-determined nor the flow of creative activity be rigidly scripted (Scardamalia & Bereiter, 2014b).

Collaborative design is an essential aspect of the invention and making processes. In accordance with the Learning through Collaborative Designing (LCD) model (Seitamaa-Hakkarainen & Hakkarainen, 2001), we examine co-design as communal efforts of creating artifacts by solving complex and ill-defined problems through iterative processes in which design ideas are elaborated and refined through analysis, evaluation, deliberation, sketching, prototyping, and making. Previous studies of knowledge creation processes suggest that advanced collaboration requires group members to focus on a shared object that they jointly construct during the design process (Barron, 2003; Hennessy & Murphy, 1999; Kangas, Seitamaa-Hakkarainen, & Hakkarainen, 2013; Paavola & Hakkarainen, 2014; Seitamaa-Hakkarainen et al., 2001). The knowledge-creation process may be seen to be guided and directed by envisioned “epistemic objects” that are incomplete, being constantly further defined and instantiated in a series of successively more refined visualizations, prototypes, and design artifacts (Ewenstein & Whyte, 2009; Knorr-Cetina, 2001). Inventions can be designed only through repeated iterative efforts of solving complex problems, overcoming obstacles and repeated failures with practical experimenting, obtaining peer and expert feedback, trying again, and ending up with outcomes that may not have been anticipated in the beginning. The important aspects of craft making are “designerly” ways of thinking and the manipulation of materials and tools (see Cross, 2006). The importance of participating in embodied design activities, and working with concrete artifacts, in learning has been emphasized by many researchers (e.g. Blikstein, 2013; Kafai, 1996; Kangas et al., 2013; Kolodner, 2002). Making is a very effective way of engaging students in a design mode (Bereiter & Scardamalia, 2003) that guides them to focus on the usefulness and adequacy of ideas and, moreover, invest efforts in the continuous improvement of design ideas.

The basic tenet of our investigation is that every student could be an inventor, regardless of gender, school achievements, ethnicity, or other characteristics. The co-invention projects are intended to provide diverse students a strong sense of contribution, that is, they experience that they are doing something worthwhile together, each student’s unique efforts and accomplishments matter, and that the whole team is jointly reaching something that no one could have done on his or her own. The success of collaborative creation of knowledge is critically dependent on students who actively engage in and take responsibility for the process (Paavola & Hakkarainen, 2014; Sawyer, 2006; Scardamalia & Bereiter, 2014a). Focused creative pursuit requires students to actively work toward a joint epistemic object, to listen, understand and help each other, and to engage in shared efforts of testing and constructing artifacts being developed (see e.g. Barron, 2003). The present type of technology-enhanced making processes has not been formerly implemented in basic education. Co-invention projects may be experienced as very challenging by students and their teams because of working with unfamiliar digital fabrication technologies, encountering unanticipated construction problems, and carrying out design inquiries leading to unforeseen directions (see Zhang, Scardamalia, Reeve, & Messina, 2009). These processes may overwhelm the students, if they are not supported adequately (e.g. Linn, 2006). It is essential to understand, how students participate and collaborate in a small group setting when participating in open-ended co-design and making processes, if we want to design successful pedagogical approaches and practices that advance knowledge-creating learning. Collaboration in small teams of students has been investigated quite rigorously, especially from the perspectives of collaborative talk and actions (see e.g. Ching & Kafai, 2008; Hennessy & Murphy, 1999; Kangas et al., 2007, 2013; Linn, 2006). However, how open-ended co-design and co-invention processes, aimed at the creation of new complex artifacts, truly unfold through student teams actually organizing and working from phase to phase throughout the process calls for further investigation (cf. Hennessy & Murphy, 1999). This present study seeks to contribute especially to this area of research.

Research setting: Co-invention project in Aurinkolahti basic school
The present investigators organized a co-invention project with the Aurinkolahti basic school in spring 2017. All of the Grade 7 classes, 70 students in total, aged 13 to 14, participated in the project. Finnish curriculum for basic education involves compulsory weekly craft lessons until Grade 7. Integrated design and making activities, which are characteristic of Finnish craft education, provide ample opportunities for bringing together STEAM subjects. This enabled us to implement learning-by-making projects as a part of regular curricular activity. For assistance, teachers relied on collegial (co-teaching) resources to negotiate emerging challenges: we engaged two craft subject teachers with three other subject teachers (science, ICT, and visual arts) to orchestrate the project. Moreover, we engaged Grade 8 students to function as “digital technology” tutors to provide additional support in guiding the
student participants. The Innokas network (innokas.fi/en) offered support regarding digital instruments, materials, and coding initially to the tutor students and when necessary also to the inventor teams. The teachers were familiarized with the technologies used as well as given pedagogical support.

The co-invention challenge, co-configured between teachers and researchers, was open-ended: “Invent a smart product or a smart garment by relying on traditional and digital fabrication technologies, such as GoGo Board, other programmable devices or 3D CAD”. Before the project, the Grade 8 tutor students arranged a GoGo Board workshop for every participating class, so as to familiarize the students with the possibilities and infrastructure of the GoGo Board and to promote the emergence of ideas on how programmable devices could be utilized in the inventions (cf. Ching & Kafai, 2008). GoGo Board is an open-source hardware device, developed at the MIT Media Lab, for prototyping, educational robotics, scientific experiments, and environmental sensing (Sipitakiat, Blikstein, & Cavallo, 2004). The actual co-invention project was initiated in February, with a two-hour ideation session, arranged in collaboration with the Finnish Association of Design Learning. During this session, the students self-organized into teams and constructed the preliminary ideas of their inventions. The project involved eight to nine weekly co-design sessions (two to three hours per session) during March, April and May 2017. The teams also presented their co-inventions in two Co4-Lab events, held at the University of Helsinki.

Research aims and methods
By relying on ethnographic video data and participant observations of the student teams’ co-invention processes, the present investigation focuses on examining the student teams’ participation in co-design and making processes. The specific research questions guiding our investigations were as follows: 1) What was the nature of the co-design processes of the student teams? 2) How did the student teams organize and collaborate during the co-design processes? 3) What kind of co-design-related differences occurred between the teams? 4) How did the co-design approaches and the nature of collaboration relate to the success of the co-invention processes?

For video recording and intensive follow-up by the first author, we randomly selected two whole classes that consisted of seven students’ co-invention teams. The recordings were carried out separately for each team, using an individual GoPro action camcorder, placed on a floor-standing tripod, and a separate wireless lavaliere microphone for each team. The camera was positioned at a high side angle to capture a team’s actions as fully as possible. The first author was also present during every co-design session and made observations and field notes to support in-depth analysis of the data. Five of the seven videoed teams were selected for the detailed analysis. One team was discarded due to fragmented data, caused by technical difficulties during the recording process. The other team was discarded because of ethical issues within the team. Diverse invention projects and digital tools and fabrication methods employed provided us with very rich data (see Table 1).

<table>
<thead>
<tr>
<th>Name</th>
<th>Members</th>
<th>Data</th>
<th>Basic ideas of the inventions</th>
<th>Digital technologies used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike</td>
<td>3 boys</td>
<td>14:07</td>
<td>A three-wheel bike that contains smart technologies, such as an environment responsive, rechargeable LED lighting system</td>
<td>GoGo Board</td>
</tr>
<tr>
<td>MGG</td>
<td>4 boys</td>
<td>13:15</td>
<td>MGG (Mobile Gaming Grip), a pair of handles that improves the ergonomics of a mobile phone while playing games</td>
<td>3D CAD modeling, 3D printing</td>
</tr>
<tr>
<td>Moon</td>
<td>6 girls</td>
<td>13:09</td>
<td>A smart outfit for sports, including for example. an environment-responsive lighting system to improve safety</td>
<td>Adafruit Flora and Gemma, light sensors, RGB LEDs</td>
</tr>
<tr>
<td>UrPo</td>
<td>6 boys</td>
<td>12:34</td>
<td>A smart outsole for sport shoes, including for example an automatic warming system for winter sports</td>
<td>Adafruit Flora and Gemma, temperature sensors</td>
</tr>
<tr>
<td>Plant</td>
<td>7 girls</td>
<td>12:21</td>
<td>An automatic plant care system, that incorporates decorative elements.</td>
<td>GoGo Board</td>
</tr>
</tbody>
</table>

A theory-driven coding template was constructed for characterizing the co-design and making processes. The coding template and codes, relevant to this research, are presented in Appendix A. Codes for verbal and embodied design actions were based on design research literature and the Learning by Collaborative Design (LCD) model (e.g. Goel, 1995; Seitamaa-Hakkarainen et al., 2001; Seitamaa-Hakkarainen, Viilo, & Hakkarainen, 2010), as well as on our earlier experiences of investigating maker-centered learning (e.g. Kangas et al., 2013; Lahti et al., 2016). The data analysis was based on systematic observation and coding of the videos. The videos were coded in 3-minute segments using the ELAN multimedia annotator (4.9.4 and 5.0.0-beta). For every segment, primary design actions were determined for the whole team and possible subgroups of students. Instances of design process engagement were emphasized over non-task-related actions and speech. For instance, if a team discussed non-task-related issues while being actively making a prototype, the segment was coded as model making rather than non-task-related action. The present analysis aimed at providing an overview of the co-design processes; in
subsequent studies, we will zoom into more detailed microlevel analyses of the data. Four investigators took part in coding the data. After the coding process, all segments containing no action were removed from the video data. To avoid distortion of the data, the data was further adjusted by removing intervals of process organizing or non-task-related actions, such as when a team had to wait and was therefore unable to continue the design process further (e.g. waiting for a teacher to arrive to give obligatory instructions for the safe use of tools, or waiting for computers to open or update). The final, adjusted video data, consisted of 65 hours, 27 minutes of coded team session videos in total, 12 to 14 hours per team (Table 1).

To make the massive amount of complex data from the teams’ and individual students’ processes analyzable, we developed a new method for visual analysis of the data. Based on verbal and embodied design actions of the teams and individual students, systematic process visualizations, striped process rugs, in the form of color-coded, layered diagrams, were built. The striped process rugs, presented in Figure 1, represent the teams’ design processes as design actions in 3-minute stripes. Every design session of a team is separated with a blank horizontal stripe, with the first session being on top. Blank areas in participants’ individual design processes signify that the participant was absent. In stripes where both verbal and embodied design actions occur simultaneously, verbal design action is presented on the left side of the segment and the embodied design action on the right. The construction of the striped process rugs not only provided us with insight into the overall design processes of the teams and activities of individual team members, but also preserved plenty of information about collaboration within the teams. When necessary, results drawn from the process rugs were compared to the ethnographic observations, which were made by the first author, or to the corresponding sections of the raw video data to further verify the findings or to gain further information.

Nature of the student teams’ co-design processes

In terms of the process outcomes, four teams (Bike, MGG, Moon and UrPo) achieved a successful co-design process. They developed well-articulated design ideas, produced visualizations and prototypes, and tested and refined their design ideas. On the other hand, the Plant team failed almost completely on all the above outcomes. They produced only a few separate objects that did not have any functionality. The process rugs indicate clearly that the co-design processes varied significantly between the successful teams and the Plant team. The successful processes were iterative, yet still progressing, in nature. The successful teams spent most of the working time deeply engaged in the design-related activities, especially model making and digital experimenting. The process rugs indicated that the periods of design actions appear to be longer and more coherent for the smaller teams (i.e. Bike and MGG) than for the larger teams. Furthermore, the successful teams, especially the smaller ones, spent very little time on non-task-related activities, whereas the Plant team spent most of their working time on non-task-related actions, to an extent that some sessions were spent almost entirely doing non-task-related things. In addition, the Plant team’s co-design process was very scattered in nature. The lack of design actions, especially model making, compared with the successful teams was remarkable. The design actions varied between the teams also due to the differences between the inventions and fabrication methods. In the following paragraphs, we will describe each teams’ co-design projects.

The invention of the Bike team was a three-wheel bike that contains an environment-responsive, rechargeable LED lighting system, utilizing the GoGo Board. During the first project sessions, they made mechanical experiments for possible structures of their bike. Based on their experiments and the knowledge they sought from the Internet, they refined their ideas intensively, especially during the second working session. Their co-design process was clearly iterative in nature in terms of stages of making the actual model and further refining the design idea alternating. Towards the end of the project, they crystalized their idea and concentrated mostly on the model making. They used several advanced fabrication methods that were all new to them, such as welding and metal lathe turning. Simultaneously, they had to learn the new methods of fabrication, consider the final product and its mechanics, deliberate on materials and structures, as well as, organize their process.

The MGG team invented a mobile gaming grip, a pair of handles that improve the ergonomics of mobile phones in gaming contexts. They had a two-stages process, in the sense that they first built a concrete prototype from basic materials (wood, rubber and masking tape), and then, from session six onwards, created 3D CAD-models based on the first prototype. When building the first prototype, they worked very iteratively, ideating, testing their ideas, evaluating and refining the ideas further. They paid particular attention to the ergonomics and usability of the handles and on how to make the handles suitable for as many smart phones as possible. Their process highlights the importance of embodied and concrete model making, although the final fabrication method was a 3D CAD model, and later a 3D-printed model. The second phase, the 3D CAD model making began with a session during which they tried to create the 3D-model with SketchUp. However, that turned out to be too complex for the team. Therefore, they decided to spend the next session experimenting with three different 3D modelling software options (Blender, SketchUp and Tinkercad). Based on the experimenting session, they chose...
to use Tinkercad and SketchUp together for the modelling, and at the end, they were able to produce a printer-ready 3D model of the handles.

The Moon team stepped into the world of e-textiles in their design process and invented an environment-responsive outfit for sports (cf. Litts et al., 2017). During the first sessions, the team concentrated on crystallizing their idea and on pattern making for the clothes. Towards the end of the process, the team engaged in three separate but partially interlinked activities: sewing the actual clothes, programming and assembling the electronics, and making presentation material for their product. During their process, they designed and made the actual clothes from elastic materials, designed the electronic circuits, sewed the conductive threads and programmed the electronics for the clothes— all actions that they had never done before. They used Adafruit Flora, a wearable electronic platform, lighting sensors, and programmable NeoPixel LEDs as their electronic components.

The UrPo team invented a smart outsole for sports shoes that can also be regarded as an e-textile (cf. Litts et al., 2017). They used Adafruit Flora and Gemma as the electronic platforms to produce the functionalities

Figure 1. Striped process rugs. P = teams’ primary design actions, 1…n = actions of individual team members.
of the insole. Additional challenge emerged from the temperature sensor-controlled warming system that they designed for the insole; they had to design this functionality from scratch, using resistance wire. Their design process became a truly iterative one, where the refining of ideas occurred even during the second-to-last session. The team produced several concrete prototypes and sketches of different and alternative structures of the insole, especially elaborating on the placement of the Adafruit Gemma board on the insole.

The Plant group intended to build a plant care system that also served as a decor element. However, their process was very scattered in nature and did not lead to any actual prototypes. They only produced a few sketches and simple objects that could have been used to build a prototype. Sketching, which mostly meant producing colourful drawings without any new contribution to the idea, was their most prominent design practice. They were guided to make tests with the GoGo Board and later with possible power supply and connection of a pump system, but they did not fully engage in these activities and mostly left the work to the Grade 8 peer tutor students.

Collaboration and process organizing in the co-design processes

From the process rugs, the differences between the collaboration of the teams emerge clearly. The two smaller teams, Bike and MGG, worked through the whole process in intensive, jointly organized collaboration. The Moon and UrPo teams often worked together, but divided tasks among the team members more often than the two smaller teams. Especially in the larger teams, some students were more active and orchestrated the process more than the other team members. Nevertheless, they did not dictate the processes at any level. In the Bike and MGG teams, all participants were highly engaged in the processes and participated. Even when some of the students did non-task-related things, their engagement with the process did not deteriorate. For example, during MGG teams’ fourth session, student number 2 modified the fine details of the first prototype. During the modifications, the other students did non-task-related things around the table, but every time a modification was made, the whole group rapidly gathered to evaluate the modifications and to decide upon, how the handles should be further improved.

In the Moon and UrPo teams, the slight scattering of the collaboration demonstrated itself during periods of ideation, evaluation and refining of ideas – the most important phases of decision-making. Although the teams, in most cases, gathered together to make decisions, periods of divided attention occurred more often than in the Bike and MGG teams, during the decision-making activities. Further scattering also occurred, from time to time, in the larger teams, due to some students drifting to non-task-related activities, although they could have engaged in the co-design process and collaboration. The Plant team’s inability to collaborate may be because the dominant figures within the team, students 2 and 3, were also the ones who concentrated mostly to non-task-related activities. The team tried to organize from time to time to start working on the task, but often this did not lead to any actual productive design actions. The relatively long time spent on process organizing is striking on the process rugs. In addition, it must be noted that the actions related to the design process, during the last four sessions, mostly occurred when the teacher or one of the tutor students was present in the group. None of the group members took responsibility for the process. In contrast with the Plant team, the process rugs indicate that the process organizing in the successful teams often led to distinct design actions. The organizing occurred typically at the beginning and end of the sessions and when students or the whole team moved from one design activity to another. The process rugs highlight how the larger teams, although successful, required more time to organize compared with the smaller ones. Furthermore, the larger teams needed more guidance on how to organize their processes than the Bike and MGG teams did.

Both the success and collaboration within teams appeared to relate to the team sharing the same epistemic target object. It was very clear from the videos and the observations made that the Bike and MGG teams shared the same epistemic object throughout the whole process that they actively sought to develop. Also, the Moon team, although their collaboration scattered every now and then, seemed to share the same epistemic object across the design process. The UrPo team, however, seemed to concentrate on different epistemic target objects, especially during sessions three and four, although they regrouped around one object towards the end of the project. In contrast with the successful teams, the Plant team did not have a shared epistemic target object at all. They had a shared idea at the beginning, but it did not evolve or crystallize in any way.

Conclusions and discussion

The purpose of the present investigation was to analyze five student teams’ co-design processes that involved using traditional and digital fabrication technologies for inventing and making complex artifacts. The systematic coding procedure and visualization of the group processes, based on that coding, proved to be very effective methods to analyze the large and complex group session video data. The striped process rugs provide deep insight into the processes of the co-invention teams and the individual students. The amount of information that such visualizations can convey, at a glance, would be very difficult, if not impossible, to produce using other analytic
methods. When supported with the observations made during the co-design sessions by the researcher, and the ability to return to the raw video data, the results can be, both, confirmed and deepened. The 3-minutes analysis segment size is, however, likely to have reduced the occurrence of certain activities in the analysis results. For example, short moments of reflection or ideation may have been overridden by more prominent, longer lasting actions. However, we managed to disclose the iterative nature of the successful co-design processes, using the 3-minutes analysis segment. The process not only provided us with the results of this research, but also offer us information on those parts of the data that could be interesting grounds for further analysis from other perspectives, such as more detailed micro-level analyses on how conceptual and materially embodied aspects of knowledge-creation inter-related during the processes.

A very promising finding of this research is that most of the teams were able to participate successfully in the co-invention project, although coming up with successful solutions required overcoming both epistemic and technological challenges. The four successful teams were able to design and invent complex, targeted artifacts. The importance of a shared epistemic object of the design process and collaboration was prominent on every team. Our findings further confirm the results of previous studies made about the importance of such objects for co-design and knowledge-creation processes (Paavola & Hakkarainen, 2014; Seitamaa-Hakkarainen & Hakkarainen, 2001). We also suggest, in agreement with previous research, that the importance of concrete making cannot be over-emphasized in these types of co-invention processes (cf. Kafai, 1996; Kangas et al., 2007, 2013; Kolodner, 2002). We argue that without designing and making the concrete prototypes, the processes of every team would have lacked key opportunities for improving the students’ ideas, which is a critical aspect of knowledge creation.

In addition, we must commend the importance of the crafts subject teachers and the Grade 8 tutor students for the successful completion of the project; their guidance enabled student participation in the advanced design processes and concrete prototyping. Teacher expertise regarding design, fabrication methods, mechanics and materials, as well as on the pedagogics of invention and making appear to be crucial when conducting these types of knowledge-creating projects. The present results indicate further that team size has a significant effect on the nature of the team collaboration. While the two smaller teams (Bike and MGG) worked throughout the whole process in very intensive and close collaboration, the larger teams (Moon, UrPo and Plant) scattered, at least to some extent. The collaboration in the small teams was also more democratic and balanced than in the larger teams. However, to confirm these findings, more research into similar co-design processes is needed. When continuing these design experiments, we will limit the team size.

The present investigation reveals that significant aspects of the maker culture can be integrated with the regular curricular activity of Finnish schools. By relying on traditional and digital fabrication technologies and practices, young students can be engaged in co-designing, co-inventing, and joint making of complex artifacts, sparking intellectual, technical and aesthetic challenges. Students who have completed successful co-inventions projects, such as those of the Bike, MGG and Moon teams, will be engaged as peer tutors for the new cohort of student inventors regarding programmable devices, 3D CAD-modeling and 3D-printing. Rigorous research as well as collaboration with teachers, students, and other players in the field of education is needed for expanding maker-centered learning at school.

References


Acknowledgments
Academy of Finland grant 286837 (Co4-Lab) and Strategic Research Council grant 312527 (Growing Mind).

Appendix A: Structure of the coding template and code descriptions
Description and notes (written description of what the group was doing and notes for later analysis)

Primary verbal design action: (“Non-task-related action” code was used only if there was no applicable primary embodied action.) Codes: seeking information, process organizing, analysis, ideation, evaluation, refining idea, describing, discussion about manufacturing, non-task-related action, no action

Primary embodied design action Codes: drawing/sketching, material experimentation, mechanical experimenting, digital experimenting, making presentation material, model making

Pupil 1…n: (Left empty if the pupil is absent.) Codes: active, passive

Group work: Codes: all together, divided. For divided group work three additional codes were applied: subgroup 1…n (including student numbers divided by comma, e.g. Subgroup 1: 1,3 and Subgroup 2: 2), subgroup 1…n verbal design action (See primary verbal design action for codes, applied individually for every subgroup.), subgroup 1…n embodied design action (See primary embodied design action for codes.)
Trade-offs in Using Mobile Tools to Promote Action with Socioscientific Issues

Susan A. Yoon, Jooeun Shim, and Noora Noushad, yoonsa@upenn.edu, jshim@gse.upenn.edu, noora@gse.upenn.edu
University of Pennsylvania

Abstract: In this study, we build on an emerging trend in socioscientific issues (SSI) education to support action through the use of personally relevant digital tools. We investigate the design of curriculum that integrates SSIs with the programming and construction of mobile apps. Through a series of design iterations, we highlight important tradeoffs in design choices that can potentially impact the depth of students’ learning of SSIs and how students take action. These considerations include the sequencing of programming versus SSI instruction, enabling or restricting student choice of SSI topics, mandating collaboration on app development, and emphases on packaged computational components versus computational concepts. We conclude with several design suggestions to maximize efforts to promote action through app construction and SSI education.

Introduction
When promoting scientific action among students, it is important to provide context for knowledge (Buxton, 2010), motivation to learn the science content (Skamp et al., 2004), and tools to transfer the knowledge into action (McNeill & Vaughn, 2012). However, it can be difficult to design learning environments that engage learners in authentic investigation of science content while creating motivating tasks that will encourage action without compromising learning goals (Alchin et al., 2014). In this study we tackle the challenge of providing a context, motivation for learning, and means to transfer knowledge into action among middle school students by creating an environment that integrates socioscientific issues (SSIs) with the construction of mobile apps. Here, we provide a lens to understand educational challenges by describing the tradeoffs that we encountered in designing such a learning environment. Previous research has suggested that designers need to discuss the affordances and constraints unique to design iterations so that educators and curriculum designers can be informed about important changes and rationales for these changes that take place (Peppler et al., 2016).

Conceptual framework
We draw on three areas of scholarship to support the study goals—the use of SSIs in science education (Barton & Tan, 2008), mobile learning (Price et al., 2014; Sharples & Pea, 2014), and the programming of digital learning tools (Resnik et al., 2009; Werner, Denner & Campe, 2012).

Socioscientific issues
Teaching science through SSIs represents what we know best about how people learn. It makes science personally relevant (Karahan & Roehrig, 2015); enables the use and practice of domain specific skills such as scientific reasoning and argumentation (Sadler, 2004); and enables awareness of complex scientific issues that impact social and environmental conditions from multiple perspectives (Burek & Zeidler, 2015). The notion of action has also been highlighted as an important goal for SSI instruction (Lee, 2015). This moves students out of the place of “arm chair critic” (Hodson, 2003) to a place where they can work to improve the communities they live in. However, action is challenging to enact as students are not exposed to ways in which they can apply their reasoning skills to real world activity and are given limited opportunities to demonstrate their content knowledge beyond capstone presentations and classroom debates (McNeill & Vaughn, 2012). A recent content analysis of 122 SSI-themed studies in the top 5 science education journals found that less than 2% were focused on engaging in citizenship work which can be arguably interpreted as an action orientation (Tekin et al., 2016). To address this issue, we build on an emerging trend in SSI education that enables students to create personally relevant tools (Karahan & Roehrig, 2015) to provide them with mechanisms through which action can be taken.

Mobile learning
The recent trends in mobile learning have also revealed mechanisms for students to take action. The unique affordances of mobile learning platforms such as the ability to move to different locations along with technology-enabled affordances like location-aware sensors have been leveraged to engage learners in contextualized learning activities (McQueen et al., 2012; Sharples & Pea, 2014). Consequently, research in using mobile apps to support learning has shown promising results in promoting student engagement and motivation in STEM (Grover & Pea,
2013; Ni et al., 2016; Price et al., 2014). However, many mobile learning initiatives have tended to put the learner at the user end of mobile app engagement rather than allowing learners to construct the apps themselves (Kearney et al., 2012). Under this learning scenario, only the designers have real agency to create functions for specific purposes whereas users have little agency over function or purpose (Burrell, 2016). A specific goal of this study was to engage students in designing mobile apps that could serve a chosen function for a chosen purpose with the hope of encouraging student agency, and in turn, student action.

Programming digital artifacts
Relatedly, designing and then developing the mobile app requires programming ability. Programming can help to develop confidence in learners to deal with complex phenomena, work through challenging problems, and promote the setting and achieving of goals (Barr & Stephenson, 2011). Furthermore, creating digital artifacts has been shown to develop learners’ reasoning skills while simultaneously embedding knowledge in relevant cultural and personal activity (Resnik et al., 2009; Werner, et al., 2012). With the advent of blocks-based programming platforms such as Scratch (Resnick et al., 2009), novice programmers are supported with computational logic that is built into the programming environment thereby eliminating frustrating syntax errors. This has, in part, influenced the marked increase in interest in helping all students become programming literate, e.g., Hour of Code; Girls Who Code, over the last several years. Yet, a major challenge exists in terms of finding authentic ways to integrate computer science with more mainstream subjects such as Science and Math in order to contextualize and make computer programming more relevant (Sengupta et al., 2013).

Solving local socioscientific problems by constructing mobile apps
In this study, we aim to address the challenges in the aforementioned literature through the design and construction of mobile apps that investigate a local socioscientific issue that has import in the students’ local environment. We use the mobile learning platform called App Inventor to support these local investigations. The App Inventor platform uses a graphical programming language similar to Scratch in which computational procedures are built into easily assembled blocks (http://appinventor.mit.edu/explore/front.html). Grover and Pea (2013) highlight several benefits of using tools like App Inventor to promote computational thinking, interest, and access. They write that such tools are underpinned by the principle of “low floor, high ceiling,” which means that the environment has a low threshold for learning the initial programming language but embeds opportunities for more advanced computational investigations.

The broad goal of the project was to understand the extent to which building mobile apps with an SSI focus could motivate students toward scientific action with content specifically anchored in science. Through two design iterations, we found several important trade-offs to consider in the design of curricular activities that appeared to have an impact on student learning and participation outcomes. In this paper, we first describe the design of our curricular activities in the two design iterations that encompassed a spring and a fall elective class with 7th grade students. We discuss changes that were made in each class’s design in order to improve student learning and participation outcomes and we describe the tradeoffs that similarly-minded designers should consider when developing learning programs with these educational goals.

Methods
Design and intervention
In this exploratory study, we use a design-based research (DBR) methodology. DBR studies require interventions to run through cycles of conceptualization, design, implementation, evaluation, and redesign until results show promising outcomes in learning measures (Puntambekar & Sandoval, 2009). Both elective courses ran twice a week for 45 minutes each day over a 12-week cycle. The curriculum was delivered in 3 blocks. In the first iteration, block 1 focused on helping students learn the App Inventor programming language. Students were given tours of code, asked to create mini-apps such as how to make sounds and how to create an action by shaking the mobile tablet, and introduced to app cards that taught students more nuanced programming functions. During block 2, students explored the meaning of socioscientific issues first by learning about global challenges such as climate change, hurricanes, and the overabundance of garbage. They were then asked to think about their local community and issues related to science that they could examine. They brainstormed issues that they wanted to solve or that they could relate to, and that could be amenable to integration in an app. Next, in teams of two, students built paper prototypes and constructed their apps. Finally, in block 3, students tested and revised their apps. Based on results from the first iteration, several design changes were made in the second iteration. First, the topic of SSIs was presented before any App Inventor programming instruction. We differentiated roles between teams of
students to focus either on the programming or on the science. And we limited the number of programming ideas that students could use in the construction of the app.

Participants
We recruited 25 seventh-grade students who chose to participate in our study as an elective course in the spring and fall of 2016 from a public school located in a large urban school district in the north east part of the United States. In the spring semester class, 13 students (6 female and 7 males) participated and in the fall semester class, 12 students (4 females and 8 males) participate. Across the two groups, 7 students identified as White, 7 students identified as mixed race, 5 students identified as African American, 3 students identified as Asian, and 3 students identified as Other.

Data sources and analyses
Data collected in the study included a pre- and post-intervention survey with questions that asked about students’ knowledge of socioscientific issues and programming, and interests in the application of science and technology. The survey included 10 Likert-scale questions with open-ended questions added for students to explain their ratings. Questions included:

- Do you think science is useful in your everyday life?
- Do you think learning science helps you to take action in solving problems in your community?
- Do you think learning how to make apps helps you to take action in the community?
- Do you think you will use the mobile app you have developed?

Students were also given a 20-minute post-intervention interview that probed their ideas related to the research goals as well as students opinions about how the class was structured. There were 13 semi-structured questions with multiple subquestions. Interview questions included:

- Tell me about your experience building your final app.
  - What does your app do? What motivated you to choose the topic of your final app?
- Can you tell me what socioscientific issues are?
  - Is the problem that your app solves a socioscientific issue?
  - Has building the app helped you see why these issues are important in your life? How?
- What would you have changed in the way the class was taught?

In addition, the 14 apps that were constructed in teams over the two iterations were analyzed to understand what students had created, what they had learned from the programming activities, and how they applied it in their project. Observation field notes were also kept throughout the course. These observations were focused on understanding the extent to which students were engaged and experiencing challenges both in content and interpersonal interactions. Since the methodology was exploratory and design based in nature, all data sources for this study were mined qualitatively and discussed among the project team.

Results
In general, we found that students were engaged in both iterations. The majority of students were able to produce a working app but we found that what kind of app they produced and the apps ability to function in terms of our research goals varied between iterations. We also found that student interests and knowledge of SSIs and programming were different. In Table 1, we list the apps that students constructed with a brief description of what the apps were constructed to do. We follow the table with a detailed discussion of the design challenges and associated tradeoffs that are hypothesized to have impact on the study goals.

<table>
<thead>
<tr>
<th>Team</th>
<th>App project description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>To reduce the chance of being bullied, the app gives a diet plan along with links to workout videos.</td>
</tr>
<tr>
<td>1-2</td>
<td>This app allows users to view the breakfast and lunch menu served at school and displays the nutritional value of selected items.</td>
</tr>
<tr>
<td>1-3</td>
<td>To reduce distraction while doing homework, this app times how long the user has stayed focused.</td>
</tr>
<tr>
<td>1-4</td>
<td>To help reduce electricity consumption, the user can record the time spent on various appliances.</td>
</tr>
<tr>
<td>1-5</td>
<td>To help students with their imagination, the app allows users to summon various mythical creatures.</td>
</tr>
</tbody>
</table>
like just pure CO$_2$. I knew like fizzy drinks had cause that is what they are—carbonated drinks, but I didn’t know foods exhibit a carbon impact, one student commented, “I didn’t know cottage cheese had carbon dioxide, to eat. Survey and interview responses from students in the second iteration also showed a deeper understanding of information could be crowd sourced, shared, and used by community members to make choices about where to
designed an app that identified high or low carbon packaging used by local restaurants (2-6 Table 1). This
social, scientific in nature, and included some action that could have local impact. For example, one group
idiosyncratic student-chosen topics. In the end, all artifacts addressed issues of environmental impact which were
that we could scaffold instruction with essential scientific information rather than having to address content in
of the SSI content. We also limited the topic to investigating one’s carbon footprint in the local environment so
coherent and accurate definition that included knowledge of SSIs as social issues with scientific content and that
to the survey question, “what are socioscientific issues?” only four students out of 13 were able to provide a
majority of students said that their interest in science and technology improved from participating in the course.
SSIs could be found in local contexts and addressed through local activity. Nevertheless, in their interviews, the
Based on these results, we wanted to see if we could improve on student understanding of SSIs by
programming to students in the first block before instruction about SSIs proved to be very motivating to students, however, this happened at the expense of students’ learning what SSIs were and what constituted local SSIs that they could help to improve. Despite continual redirection away from tinkering with the code, students remained disengaged from the science content in the second block where instruction of SSIs occurred. This resulted in student teams choosing to make apps that had little to no local SSI content. For example, one team created The Distraction App (1-3 Table 1) that monitored how much time they spent surfing the internet rather than doing their homework. If they managed to stay on task for a period of time, an emoji appeared to congratulate their effort. Another team created an app that would enhance imagination through an exploration of unicorns and griffins—The Magical Creatures App (1-5 Table 1). While there appeared to be some personal relevance embedded in these artifacts, there was little that could be said to encourage action in a local SSI. Furthermore, in response to the survey question, “what are socioscientific issues?” only four students out of 13 were able to provide a coherent and accurate definition that included knowledge of SSIs as social issues with scientific content and that SSIs could be found in local contexts and addressed through local activity. Nevertheless, in their interviews, the majority of students said that their interest in science and technology improved from participating in the course.

Introducing programming to students in the first block before instruction about SSIs proved to be very motivating to students, however, this happened at the expense of students’ learning what SSIs were and what constituted local SSIs that they could help to improve. Despite continual redirection away from tinkering with the code, students remained disengaged from the science content in the second block where instruction of SSIs occurred. This resulted in student teams choosing to make apps that had little to no local SSI content. For example, one team created The Distraction App (1-3 Table 1) that monitored how much time they spent surfing the internet rather than doing their homework. If they managed to stay on task for a period of time, an emoji appeared to congratulate their effort. Another team created an app that would enhance imagination through an exploration of unicorns and griffins—The Magical Creatures App (1-5 Table 1). While there appeared to be some personal relevance embedded in these artifacts, there was little that could be said to encourage action in a local SSI. Furthermore, in response to the survey question, “what are socioscientific issues?” only four students out of 13 were able to provide a coherent and accurate definition that included knowledge of SSIs as social issues with scientific content and that SSIs could be found in local contexts and addressed through local activity. Nevertheless, in their interviews, the majority of students said that their interest in science and technology improved from participating in the course.

Based on these results, we wanted to see if we could improve on student understanding of SSIs by
manipulating the design of the course. Rather than beginning with the App Inventor programming activities, in the second iteration, we started with a lengthy investigation of SSIs. This enabled a more in-depth investigation of the SSI content. We also limited the topic to investigating one’s carbon footprint in the local environment so that we could scaffold instruction with essential scientific information rather than having to address content in idiosyncratic student-chosen topics. In the end, all artifacts addressed issues of environmental impact which were social, scientific in nature, and included some action that could have local impact. For example, one group designed an app that identified high or low carbon packaging used by local restaurants (2-6 Table 1). This information could be crowd sourced, shared, and used by community members to make choices about where to
eat. Survey and interview responses from students in the second iteration also showed a deeper understanding of the science content and the relationship between science and human activity and the environment. On the topic of how foods exhibit a carbon impact, one student commented, “I didn’t know cottage cheese had carbon dioxide, like just pure CO$_2$. I knew like fizzy drinks had [it] cause that is what they are...carbonated drinks, but I didn’t know cottage cheese had [it]. I think it is a shelf-life thing. A lot of companies do that...they are more worried about the products than [the] environment.”

However, with the delayed introduction to the App Inventor programming activities, students in the second iteration discussed challenges to this design change. In response to the question about what they would change about how the class was taught, one student said, “Listening to more about programming rather than listening to how climate change is affecting our world.” Another student said that, “the beginning wasn’t really that fun.” This sentiment was captured as well in the observation notes during those classes focused on delivering the SSI content such as the following comment, “Greg asked me three times now, when they would start making and programming games (Day 4)”. However, despite students’ obvious interest in programming over learning about SSIs, they were able to reflect on how their actions and daily choices could make a difference for the environment. In his interview one student said:

I learned that throwing away plastic bags has an effect on the environment, because like when you throw away plastic bags they can go to a landfill and they all just sit there and it takes them

| 1-6 | To increase awareness about air/water pollutants, the app provides various dangers of air/water pollution with animated videos. |
| 1-7 | This app is a game that sorts’ random trash correctly into 'recycle', 'compost', and 'trash' bins. |
| 2-1 | This app shows where users can leave and take plastic bags and allows users to add their own locations. |
| 2-2 | To make recycling fun, the app provides a recycling game. Users can share scores and complete action goals (e.g., recycle 15 pieces of trash in a public place) to gain higher scores. |
| 2-3 | This app helps to identify high carbon footprint foods with the quiz. |
| 2-4 | This app educates users by presenting examples of ways to lower one’s carbon footprint with a quiz. |
| 2-5 | This app provides access to a link that rates products based on their carbon footprint. They can play a game in different characters, and share the app with friends. |
| 2-6 | To help users’ make better food choices at local restaurants, the app provides a quiz as well as records shareable food comments with photos. |
| 2-7 | This app gives a quiz to inform people around the world what foods have a low carbon footprint. |
years to go away so it’s not good for the environment and if they are burned they can produce a lot of like you know toxic chemicals that can pollute the air. So like my family we have this bin where we keep all the plastic bags in and I have always wondered like why my mom doesn't just throw away the plastic bags and now I see why she does that. (Post-interview, ID 7, Marco)

Interestingly, in the pre-intervention surveys, this student said that he couldn’t readily see what he was learning in school science as all that applicable to his everyday life. The data reported here provide evidence of students’ preferences for app programming over learning the science content. It is a useful tradeoff to consider especially if time is limited, or when considering what the content goals are.

Differentiating collaboration roles

Collaboration was emphasized throughout both courses. However, how teams organized their work differed across iterations. During the first iteration, students were given the freedom to decide how to distribute project tasks between themselves. We thought that by allowing students to direct their app design with little instructor monitoring, this would create a less intimidating environment for the teams so that they could choose where to apply their respective strengths and interests. However, this happened at the expense of both members equally engaging critically with programming or with SSI content. With little exception there was one member of the team who did all the programming. For example, the PAS app (1-2 Table 1) was mainly constructed by Pete. His partner Craig did not do any programming until the fifth week of the class and only did so because Pete was absent. In their respective interviews, Pete and Craig offered differing evaluations of their contributions. When asked what they saw were the benefits of working on a collaborative team, Pete said this exit interview:

I think there are pros and cons of both [working along or working together]. One of the cons [you face] is you could get a teammate who doesn’t really do much but there is nothing you can really do about that. Pros is if you have someone that is working then it can be a lot less stressful as you both contribute to a big project as you split the work 50/50.

Craig said that he preferred “working in teams, so that one person wouldn’t have to do all the work and it would be equally divided.” He described his contribution in this way. “I had figured out the nutritional value and [Pete] had programmed the screens linking it to the menu.” From observation notes, we saw that while Pete was working hard each week, Craig spent most of the time socializing and distracting others in the class.

In the second iteration we decided to formalize the contributions in such a way that the work could be perceived as more equally distributed. With the added emphasis on SSIs, we instituted roles that different members of the team could take charge of. These roles were the science driver or the coding driver. We also introduced a pair-programming rule such that the team members switched every half hour between building the app and researching the science materials. Indeed, in all the exit interviews, students said that they felt there was equal participation among team members. However, despite continued effort in the instruction to insist that every member took a turn to program, for the most part, the team members remained in their respective roles throughout the course with little switching. This led to a clear disadvantage for those students who could become more adept at programming with a little urging. For example, in his exit interview, Emmet said, “Knowledge expert, yeah I'm clever, I wouldn't say I'm smart but I'm clever, I know how to get around things. Well, so I like knowledge more than I like [programming]...[if] I find something I know I’ll never get the hang of, I'd rather someone else do it...trying to do a Scratch project once...woh”. He also mentioned that he signed up for the course because all of his friends were talking about computer science and he wanted to improve his programming skills. However, as he admitted in his interview that he was more comfortable with science content, given the choice, he decided to remain in his comfort zone rather than challenging himself in an area of lesser strength and interest.

Similarly for those students with a clear interest in the programming side, deeper level exploration of the scientific content was given short-shrift. Mike said this about the science content in his exit interview:

Well there were things that I didn’t really understand and honestly didn’t go out of my way to understand. I just felt like I said I didn’t enjoy it enough to use it at home. So there were some stuff I didn’t understand, I probably could figure it out but I think there are more people more knowledgeable than me in that area.

James, who was Mike’s partner, thought that the delineation in roles was a good design choice because as he discussed, “one of us had to be working, we couldn't work on the same thing at the same time…um but we got twice as much done.”
In these data, we see an instructional issue with no clear-cut rationale in terms of whether more or less control over how the collaboration tasks is better or will benefit participation and learning outcomes. In the first iteration there was a perception (with good reason) of an unequal distribution of work. Correcting for this in the second iteration where we defined roles with suggestions to switch, this led to, for some students, preemptively limiting the depth of understanding that could have been experienced in either of the two content areas of science and computer programming.

**Computational components versus computational concepts**

The graphical programming language in App Inventor was used as a catalyst for students to transform their scientific knowledge into actionable artifacts, i.e., mobile apps that helped to address local SSIs in both iterations of the course. However, how programming instruction was delivered differed across iterations. For clarity, we differentiate between computational components and computational concepts. In this paper we define computational components as the various features of the App Inventor interface such as embedding videos or programming data collection using sensors; and computational concepts as more universally established knowledge of computer programming such as variables, procedures, conditionals, and loops. Given that many components are prepackaged for App Inventor, in the first iteration, we made the design decision to teach about the different possible components that could be programmed into the app, for example the camera or voice recording components. We hoped that understanding App Inventor functionality would trigger ideas for app construction on an SSI topic. We found that the apps constructed in the first course varied in both functionality and the usage of App Inventor features. However, this happened at the expense of students critically engaging with the computational concepts themselves. The analysis of the programming showed that students only engaged with computational concepts at a superficial level. For example, the PAS app (1-2 Table 1) used eight types of components to create their mobile solution. In the four screens that they programmed, they used the components of ActivityStarter, Accelerometer, Sharing, Buttons, Table Arrangement, Label, Textbox, and Canvas. However, we can see in Figure 1, that no significant computational concepts were used in the app. This figure shows that in each of the five event handler blocks, they did not use variable values, conditional operations, or procedures for code organization. The blocks were simply used to open other screens or to link to a web address.

To address this issue, in the second iteration, we constructed four mini-apps based on common types of app activities (i.e., a quiz app, a game app, a memo app, and a drawing app). Through these mini-apps, we modelled how computational concepts could be applied through various App Inventor features. For example, students were asked to practice making a different quiz on a blank screen by cutting and pasting code. Learning first principles of computer programming in addition to the affordance of remixing the mini-apps resulted in the majority of apps in the second iteration showing applications of computational concepts less complex notions (e.g., variables and lists) to more sophisticated ideas (e.g. conditionals, procedures, and databases). For example, Figure 2 shows the Paper Toss Race app (2-2 Table 1) in which the computational concepts of variables, lists, procedures, conditions, Boolean logic, data, and sharing created the various functionalities of the app such as the selection of game goals, a point system, and adding items to and picking items from a list in addition to 18 types of App Inventor components.
One trade-off to note here however is a lack of diversity of the kinds of apps constructed. Most apps were modified versions of the apps that were modelled for them. For example, five out seven apps included remixed versions of the Low Carbon footprint quiz feature.

**Discussion**

Capitalizing on the growing trend in using artifacts to promote action in SSI education, our goal was to investigate design iterations of curricula that combined instruction in SSIs with constructing mobile apps to encourage local action. We analyzed student data from both iterations to look for positive and negative design features and we found that there were a number of trade-offs in each case.

In this paper, we outlined three primary trade-offs. First, we found that with a curricular model of programming first and SSI instruction second with complete student choice of SSI topic, this led to increased student engagement but decreased knowledge of the science. On the one hand, the programming first model establishes an environment where making is fun (Peppler et al., 2016) and allows for students to engage in topics that are most proximal to their interest. On the other hand, the apps that were created in the first iteration were arguably devoid of real local action in the community. Where the focus was on SSI instruction first to strengthen their science content, students exhibited less interest. This creates a conundrum for instruction as students with heightened awareness of SSI but with decreased interest are less likely to transfer the knowledge into action (Burek & Zielder, 2015).

Instituting collaboration rules also presented challenges (Werner et al., 2012). Allowing student collaborative choices to emerge in place of enforcing them through instruction resulted in unequal distribution and ownership of work among group members. In the second iteration, an environment for co-creativity emerged as students distributed the task of researching the science content versus programming the applications among the various members of the group (Lubatkin, 2001). However, while it was clear that there was more equal distribution of work, students more inclined to focus on one or the other of assigned tasks did so, which prematurely limited opportunities to learn new content and skills.

The last trade-off pertaining to how programming instruction was delivered demonstrated that when students were taught app components first without detailed description of the computational concepts embedded in the components, students showed relatively weak application of computational concepts which corroborates previous research findings (Grover & Basu, 2017). Conversely, when students were given pre-programmed model mobile mini-apps to remix, students demonstrated greater sophistication of computational concepts in their code. This might be an obvious win for the latter design choice but we also saw that the second set of apps showed much less diversity in functionality and purpose.

We highlight these design trade-offs to illustrate curricular features that will impact the desired goal of enacting action in the world through applied scientific content. We can see this information as potentially valuable for similarly minded learning scientists who may need to *a priori* establish which aspect of this action will take primacy. This is important because one consideration to note is that this instruction occurred in a course with a finite limit of about 18 hours of instruction. Overall, these trade-offs may fundamentally come down to how much choice students are given. In our study, we found that with more choice, there was greater interest but less content or skill development; and with less choice for the most part, the opposite was true.

**References**


McQueen, J., Wright, J. J., & Fox, J. A. (2012). Design and implementation of a genomics field trip program aimed at secondary school students. PLoS Computational Biology, 8(8), e1002636.


Acknowledgments
This work was funded by grants from the Gregory and E.J. Milken Strategic Faculty Support Fund and the University Research Fund at the University of Pennsylvania.
Sequencing Support for Sense Making and Perceptual Fluency with Visual Representations: Is There a Learning Progression?

Martina A. Rau, University of Wisconsin – Madison, marau@wisc.edu
Miranda Zahn, University of Wisconsin – Madison, mrzahn@wisc.edu

Abstract: We tend to assume that visuals help students learn. But visuals can impede learning if students lack representational competencies that enable them to understand what the visuals show. We consider two instructional activities that target two qualitatively different representational competencies: explicit sense-making competencies and implicit perceptual fluency. Prior cross-sectional studies showed that students’ prior knowledge determines in which sequence they should receive sense-activities and perceptual-activities. This raises the question: is there a learning progression for sense-making competencies and perceptual fluency? In this paper, we address this question with a longitudinal experiment in which 71 students worked with sense-activities and perceptual-activities integrated in six instructional units of a chemistry undergraduate course. For each unit, students either received sense-activities or perceptual-activities first. We assessed learning gains of chemistry knowledge after each unit. Results suggest a learning progression that aligns with results of prior cross-sectional studies.

Keywords: Multiple visual representations, representational competencies, explicit and implicit learning processes, sense making, perceptual fluency

Introduction

Instructors often use multiple visual representations to make content knowledge accessible to students (NRC, 2006). For example, chemistry instruction typically includes the visuals in Figure 1, which emphasize complementary concepts that students need to integrate when learning about atomic structure. Indeed, there is abundant evidence that, compared to learning with a single visual, multiple visuals can help students learn content knowledge because they allow students to form more accurate mental models (Ainsworth, 2006).

Yet, visuals can impede learning if students lack representational competencies: knowledge about how visuals show information relevant to scientific and professional practices (NRC, 2006). Students need two types of representational competencies (Rau, 2017a). Sense-making competencies allow students to explain how visuals show concepts. For example, chemistry students need to explain how the valence electrons shown in Lewis structures (Figure 1a) and inner-shell electrons (Figure 1b) have different energy levels (Figure 1c) because they reside in orbitals (Figure 1d). Students acquire these competencies via explicit processes involved in verbal explanation. Perceptual fluency allows students to quickly and effortlessly infer more than is shown in a visual. For example, chemistry students need to “mentally see” inner-shell electrons, energy levels, and orbitals based on Lewis structures. Students acquire perceptual fluency via implicit processes that are inductive and nonverbal. Analogous to Kahneman’s (2003) research on decision making, sense-making competencies correspond to deliberate System 2 thinking, whereas perceptual fluency is similar to intuitive System 1 thinking (Rau, 2017a).

Prior research showed that providing students with activities that support sense-making competencies and activities that support perceptual fluency enhances their learning of content knowledge (e.g., Rau & Wu, 2015). Further, prior cross-sectional studies showed that students’ prior content knowledge level determines which sequence of sense-making and perceptual-activities they most benefit from (Rau, 2017b). Students who had low or high prior knowledge benefited most from receiving perceptual-activities first. Students with intermediate prior knowledge benefited most from receiving sense-activities first. These findings evoke a learning progression of sense-making competencies and perceptual fluency for students at various stages of learning.
However, one cannot infer longitudinal effects from cross-sectional studies. Thus, we currently do not know how students’ benefit from sense-activities and perceptual-activities changes over the course of longer learning interventions that typically occur in courses in which instruction on a sequence of topics spans multiple weeks.

To address this question, we present a longitudinal experiment that compared sequences of sense-activities and perceptual-activities over the course of a multi-week learning intervention. We chose chemistry as a domain for this research because chemistry instruction heavily relies on multiple visuals, which poses an obstacle to students who fail to acquire key representational competencies (NRC, 2006). At a theoretical level, this experiment provides new insights into how representational competencies develop along with student’s learning of content knowledge. At a practical level, the experiment provides recommendations on how to sequence instructional activities that help students learn new content knowledge with visual representations.

Prior research

Research suggests that students need two different types of representational competencies to learn with visual representations. These competencies are acquired via qualitatively different types of learning processes. Hence, they are most effectively supported by different types of instructional activities. While we focus on cognitive learning theories, the following review illustrates the relevance of this research to sociocultural theories.

Instructional activities that support sense making with visual representations

In order to use visuals to learn new content knowledge, students need sense-making competencies: the ability to understand how different visuals show complementary information about a concept (Ainsworth, 2006; Rau, 2017a). For example, when making sense of the visuals in Figure 1, students have to explain that the Lewis structure shows valence electrons (black dots in Figure 1a), which correspond to the four electrons on the outer shell (green dots in Figure 1b), which have higher energy levels than the inner-shell electrons (green arrows in Figure 1c), and which reside in electron clouds that are larger than those of the inner-shell electrons (green shapes in Figure 1d). By making sense of such connections, students learn about domain-relevant concepts. Indeed, sense making is a key part of discipline discourse in STEM (Wertsch & Kazak, 2011). For example, chemists use visuals to discuss concepts and to solve problems (Kozma, Chin, Russell, & Marx, 2000).

Students acquire sense-making competencies by verbally explaining principles by which the visuals depict relevant information (Koedinger, Corbett, & Perfetti, 2012; Rau, 2017a). This process involves mapping visual features (e.g., dots) to abstract concepts (e.g., electron distributions) (Gentner & Markman, 1997). This allows students to distinguish relevant and irrelevant visual features and to determine which information is (or is not) shown in different visuals (Ainsworth, 2006). These sense-making processes are characterized by explicit explanations that students have to willfully engage in (Chi, Feltovitch, & Glaser, 1981; diSessa & Sherin, 2000).

Prior research yields a number of design principles for activities that enhance sense-making competencies. Such sense-activities ask students to verbally explain which visual features of different representations show corresponding or complementary information about concepts (Seufert, 2003). Further, sense-activities are most effective if they ask students to actively construct mappings (Bodemer & Faust, 2006). Finally, they should help students attend to relevant visual features (Bodemer & Faust, 2006; Stern, Aprea, & Ebner, 2003).

Instructional activities that support perceptual fluency with visual representations

Research on expertise suggests that students need a second type of representational competency: perceptual fluency, which allows them to quickly and effortlessly translate across representations (Gibson, 2000). Experts “see at a glance” what a given visual shows without any perceived mental effort (Chi et al., 1981; NRC, 2006). For example, chemists quickly see that the Lewis structure in Figure 1a shows the same atom as the shell model in Figure 1b. Experts are so efficient at extracting information from visuals that it seems intuitive. This enables them to effortlessly combine information from multiple visuals (Kellman & Massey, 2013). Cognitive psychology research attributes this high level of efficiency to perceptual chunking: visual features serve to retrieve a corresponding schema that describes related concepts (Richman, Gobet, Staszewski, & Simon, 1996). The sociocultural literature describes perceptual fluency as an important aspect of disciplinary discourse (Wertsch & Kazak, 2011). Perceptual fluency plays an important role in social interactions that involve translating among visuals. For example, Kozma and colleagues (2000) describe how experts’ fluency with visuals enables them to communicate efficiently while collaborating on complex problems. Thus, perceptual fluency allows students to infer community-specific ways of thinking beyond what visuals explicitly show (Airey & Linder, 2009).

Students learn perceptual fluency via perceptual-induction processes: they induce connections among visuals without explicit instruction but through experience with many examples (Gibson, 2000; Goldstone, Schyns, & Medin, 1997). Cognitive theories describe perceptual-induction processes as nonverbal and implicit because verbal reasoning is not necessary (Koedinger et al., 2012; Richman et al., 1996) but can interfere with
learning (Schooler, Fiore, & Brandimonte, 1997). According to sociocultural theories, community practices allow students to induce a “visual language”; students can become fluent with visuals by imitating how experts use them, without explicitly knowing what they show (Airey & Linder, 2009; Wertsch & Kazak, 2011).

Prior research yields several design principles for activities that enhance perceptual fluency. Such perceptual-activities provide many simple tasks that ask students to quickly judge what a visual shows (Kellman & Massey, 2013). Effective perceptual-activities draw attention to relevant visual features by varying irrelevant features and contrasting them to relevant features. They should encourage students to rely on perceptual intuitions about what they see rather than explaining what they see. Students should get immediate feedback only on the accuracy of their response without conceptual explanations, to engage implicit rather than explicit processes.

**Combining sense-making and perceptual-induction activities**

Prior experiments (e.g., Rau, 2017b) tested effects of sense-activities and perceptual-activities on learning of content knowledge. The experiments compared (1) a condition in which students received only sense-activities, (2) only perceptual-activities, or (3) both sense-activities and perceptual-activities to (4) a control condition that received multiple visuals but no sense-activities or perceptual-activities. Only the condition that combined sense-activities and perceptual-activities had significantly higher learning gains than the control condition.

This finding bears a theoretical question about how sense making and perceptual induction interacts in students’ learning of content knowledge. This question has practical implications for how sense-activities and perceptual-activities should be sequenced. On the one hand, explicit sense-making competencies involve knowledge about which visual features show conceptually relevant information. This may help students induce meaning from visuals when they work on perceptual-activities. Therefore, doing sense-activities first may help students learn from perceptual-activities (sense-enhancement hypothesis). On the other hand, perceptual fluency involves intuitive knowledge about how to see meaning in visuals. This may reduce cognitive load when students engage in conceptually demanding sense-activities. Therefore, doing perceptual-activities first may help students learn from sense-activities (perceptual-enhancement hypothesis).

Our prior research tested these hypotheses by comparing sequences of sense-activities and perceptual-activities. Results showed that effects depend on prior knowledge and spatial skills (Rau & Wu, 2015). Students with low prior knowledge benefitted most from receiving perceptual-activities before sense-activities. This effect reversed for students with intermediate prior knowledge, who benefitted most from a sense-perceptual sequence—especially if they had low spatial skills. The effect flipped again for students with high prior knowledge, who benefitted most from a perceptual-sense sequence.

These findings invoke a learning progression. It is possible that early in the learning process, the perceptual-sense sequence is most effective because a preliminary level of perceptual fluency helps students make sense of visuals by freeing cognitive capacity for effortful explanations. Later, the sense-perceptual sequence may be most effective because it allows students to focus on understanding how visuals show, which they can then refine in subsequent perceptual-activities that help them see this meaning automatically. Later yet, the perceptual-sense sequence may be most effective because high levels of perceptual fluency may allow students to put their implicit knowledge of visuals at the service of explicitly making sense of how visuals show concepts.

A critical limitation of our prior research is that it has not tested a learning progression. The prior studies were cross-sectional and involved relatively short interventions. That is, prior knowledge was assessed with a pretest and considered to be static for the short duration of the experiment. Hence, they did not consider how students’ knowledge level develops over the course of a longer intervention. We address this limitation with a multi-week experiment in which the same students worked through a cohesive curriculum with multiple units.

**Research questions**

We investigate which sequence of sense-activities and perceptual-activities most enhances students’ learning of chemistry knowledge (research question 1). In light of prior findings, we explore if the effect of sequence interacts with students’ prior chemistry knowledge and their spatial skills (research question 2). Further, we investigate how the effect of the sequence changes over the course of the intervention (research question 3). Extrapolating from prior cross-sectional studies, we hypothesize that for early units in the learning intervention, students show higher learning gains if they receive perceptual-activities before sense-activities. For units in the middle of the learning intervention, we hypothesize they show higher gains if they receive sense-activities first. For units at the end of the intervention, we hypothesize they show higher gains if they receive perceptual-activities first.

**Method**

To address these questions, we conducted an experiment as part of an undergraduate chemistry course.
Participants
Seventy-one undergraduate students were in the course. They were assigned to sense-activities and perceptual-activities as part of weekly homework assignments. They received course credit for completing the assignments. The instructor selected the sequence of activities to align with her course schedule.

Instructional activities
Sense-activities and perceptual-activities were implemented in an educational technology for undergraduate chemistry: Chem Tutor (Rau & Wu, 2015). Students were assigned to work on six units that covered different chemistry topics related to electron configurations and atomic structure using the visual representations in Figure 1. As all units relate to periodic table trends, the units are modular in these sense that they could be sequenced in various ways that align with different instructors’ course plans. For each unit, students first received two activities that introduced them to how the visuals show key concepts. Next, students received sense-activities and perceptual-activities, sequenced according to their experimental condition.

Sense-Activities
Figure 2 shows an example sense-activity. Students are given an orbital diagram and construct a shell model of the same atom. Then they receive fill-in-the-blank prompts to make sense of how the two visuals show atoms.

Perceptual-Activities
Figure 3 shows an example perceptual-activity. The goal of these activities is to engage students in inductive processes. To this end, students are given one visual and have to select one of four other visuals that shows the same atom. They are prompted to solve this task quickly and intuitively, without overthinking their choices and without being afraid of making mistakes. Students receive many of these one-step problems in a row. The answer choices are designed to direct students’ attention to relevant visual features by contrasting relevant visual features. The contrasting cases were selected based on learner-centered studies that investigated which visual features expert chemists pay attention to but novice chemistry students tend to confuse. If students select the wrong visual, Chem Tutor tells them they made an error and asks them to try again. This is in line with perceptual learning research that recommends providing correctness feedback only so as to encourage implicit rather than explicit processes (Kellman & Massey, 2013).
Experimental design
Students were randomly assigned to one of two experimental conditions for the entire duration of the experiment. The sense-perceptual condition received sense-activities before perceptual-activities for each of the six units. The perceptual-sense condition received perceptual-activities before sense-activities. For each unit, all students first received two regular activities that introduced how the visuals depict the chemistry concepts covered in that unit. Then, for each unit and in the sequence corresponding to their condition, they received four sense-activities and 36 perceptual-activities. We chose the number of activities based on pilot tests of how many activities yield significant gains on tests of sense-making competencies and perceptual fluency, respectively.

Measures
We assessed students’ spatial skills with the Vandenberg & Kuse mental rotations test, which is commonly used in research on spatial skills and chemistry learning (e.g., Stieff, 2013). Further, we assessed students’ knowledge of the chemistry concepts covered in Chem Tutor with pretests and posttests for each unit, developed and evaluated with chemistry experts. Two equivalent test versions were used for the pretests and posttests. For each unit the sequence in which students received the test versions was counterbalanced across pretest and posttest.

Procedure
Students were assigned to work on the Chem Tutor units as a homework assignment for their chemistry course. Students had to complete the six assignments in order to align with the course schedule. The assignments were spread across three consecutive weeks, with two units per week. Students accessed the units online. As with any homework assignment, students could take breaks as needed. For each unit, students first received a pretest, then worked on the activities as per their experimental condition, and then completed the posttest for the unit.

Results
Students were excluded from the analysis if they failed to complete the assignment by the deadline imposed by their course instructor. Of the 71 students, 62 completed units 1 and 2; 69 completed unit 3; 68 completed 4; and 63 completed units 5 and 6. For the following analyses, we consider $p.\eta^2$ as a measure of effect size, $p.\eta^2 \geq .01$ being a small effect, $p.\eta^2 \geq .06$ a medium effect, and $p.\eta^2 \geq .14$ a large effect.

Effects of sequence on learning gains
To test effects of sequence on learning gains (research question 1), we used ANCOVAs with students’ posttest scores for the given unit as dependent measure, sequence as independent factor, and pretest scores for the given unit as covariate. To test whether sequence interacts with prior knowledge (research question 2), we added interaction terms for pretest and spatial skills with sequence to the ANCOVA model. In the following, we report results from the model with the interaction effects that were significant. For significant interactions, we exami-
ined the direction of effects with effect slices that compute the effect of sequence for students with low, intermediate, and high prior knowledge, using the 33rd and 66th percentiles as cutoffs. Table 1 shows results by unit.

For unit 1, we found no effect of sequence on posttest scores, $F(1, 60) = 2.51, p = .12$, and no interaction between sequence and pretest or spatial skills ($Fs < 1$). For unit 2, we found no effect of sequence on posttest scores ($F < 1$) and no interaction between sequence and pretest ($F < 1$). There was a significant interaction of sequence with spatial skills, $F(1, 57) = 5.41, p = .01, \eta^2 = .16$. However, effect slices showed no significant effects for students with low, intermediate, or high prior knowledge ($ps > .17$).

For unit 3, we found a significant effect of sequence on posttest scores, $F(1, 66) = 6.33, p = .01, \eta^2 = .08$, such that students in the perceptual-sense condition had higher learning gains than students in the sense-perceptual condition. There was no interaction between sequence with pretest or spatial skills ($Fs < 1$).

For unit 4, we found a significant effect of sequence, $F(1, 63) = 4.15, p = .05, \eta^2 = .06$, such that students in the sense-perceptual condition had higher learning gains than students in the perceptual-sense condition. There was no interaction between sequence and pretest ($F < 1$). However, there was a marginal interaction between sequence and spatial skills, $F(1, 63) = 2.89, p = .09, \eta^2 = .04$. Effect slices showed that the advantage of the sense-perceptual condition was more pronounced for students with low spatial skills ($p = .06$) than for students with intermediate ($p = .48$) or high spatial skills ($p = .92$).

For unit 5, we found a marginal effect of sequence, $F(1, 57) = 4.15, p = .09, \eta^2 = .05$, such that students in the sense-perceptual condition had higher learning gains than students in the perceptual-sense condition. There were no interaction between sequence and pretest ($F < 1$) or spatial skills, $F(1, 57) = 2.36, p = .13$.

For unit 6, we found a significant effect of sequence on posttest scores, $F(1, 59) = 5.80, p = .02, \eta^2 = .09$, such that students in the perceptual-sense condition had higher learning gains than students in the sense-perceptual condition. There was a significant interaction with pretest, $F(1, 59) = 4.20, p = .05, \eta^2 = .07$. Effect slices showed that the advantage of the perceptual-sense sequence was more pronounced for students with low ($p = .01$) and intermediate prior knowledge ($p = .06$) than for students with high prior knowledge ($p = .76$). There was no interaction of sequence with spatial skills ($F < 1$).

Examining Table 1 allows us to answer how the effect of the sequence changed throughout the intervention (research question 3). Counter to our hypotheses, we did not see differences between conditions in early units. However, the results for units 3-6 match our hypotheses in the sense that students show a benefit for the perceptual-sense sequence in units 3 and 6 (i.e., in the first half of the intervention and at the end of the intervention) and a benefit for the sense-perceptual sequence in units 4 and 5 (i.e., in the middle of the intervention).

Table 1: Overview of sequence effects by unit. (*) indicates effects with $p < .10$; * $p < .05$; ** $p < .01$

<table>
<thead>
<tr>
<th>Unit</th>
<th>Effect on posttest</th>
<th>Direction of interaction effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Perceptual-sense &gt; Sense-perceptual **</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Sense-perceptual &gt; Perceptual-sense *</td>
<td>Especially for low spatial skills</td>
</tr>
<tr>
<td>5</td>
<td>Sense-perceptual &gt; Perceptual-sense (*)</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>Perceptual-sense &gt; Sense-perceptual *</td>
<td>Especially for low and intermediate pretest</td>
</tr>
</tbody>
</table>

Discussion

Prior research showed that students most benefit from interventions that combine activities designed to support for sense-making competencies and perceptual fluency. Further, prior cross-sectional studies showed that students’ prior knowledge seems to determine in which sequence sense-activities and perceptual-activities should be combined. These findings invoked a learning progression that had not been tested. Our findings from this longitudinal study match previous findings from cross-sectional studies. Based on our results, we tentatively put forward the following progression with, roughly, four phases. In the first phase (units 1-2), we found no difference between sequences. It is possible that at this early stage, students familiarize themselves with how the visuals show concepts, which is considered to be an important learning process that occurs before students learn to make connections among different visuals (Ainsworth, 2006; Rau, 2017a).

In a second phase (unit 3), students benefited from receiving perceptual-activities before sense-activities. This finding is consistent with the theory that perceptual fluency frees cognitive capacity that students can invest in effortful explanation-based processes while working on sense-activities. It is possible that this mechanism is important at this earlier stage because the cognitive load of explanation-based sense-activities...
would be too high and could impede students’ learning if they lack a preliminary level of perceptual “intuitions” about what the visuals show and inhibits their ability to participate in disciplinary discourse with these visuals.

In a third phase (units 4-5), students benefited from receiving sense-activities before perceptual-activities. This finding is consistent with the theory that sense-activities help students attend to relevant visual features while they work on perceptual-activities. It is possible that this mechanism is important in this middle stage because students have sufficient knowledge so that sense-activities are not so cognitively demanding that it would impede learning. Further, first focusing on how visuals show particular concepts allows students to refine their understanding of domain-relevant concepts. Subsequent perceptual-activities can then serve to refine these connections between visuals and concepts, allowing students to automatically “see” concepts in visuals.

In a fourth phase (unit 6), students again benefited from receiving perceptual-activities before sense-activities. This finding again aligns with the theory that perceptual fluency frees cognitive capacity that students can invest in effortful explanation-based processes. It is possible that this mechanism becomes important again at this late stage because now the main benefit of perceptual fluency is to put intuitive knowledge about visuals at the service of explaining how visuals show concepts—which is afforded by the perceptual-sense sequence.

Our experiment makes important novel contributions to the literature on representational competencies. We expand theory by providing new evidence that there is a learning progression of representational competencies. Specifically, our focus was on learning of content knowledge. We found that students’ acquisition of content knowledge is best supported if they receive activities that focus on implicit, then explicit, and then implicit knowledge about visuals. Our findings also have practical implications for the sequence of these instructional activities that support representational competencies in chemistry. Our findings show that if sense-activities and perceptual-activities are sequenced in a way that matches the hypothesized learning progression, they achieve higher learning gains of content knowledge.

Limitations and future directions
Future research should address a number of limitations of our study. First, our study provided chemistry content in particular sequence, which matched the schedule of the course in which we situated our experiment. Therefore, future research should test whether the findings of a learning progression remain when sense-activities and perceptual-activities are added to a different sequence of content-focused instruction. Second, while we consider the fact that our experiment was embedded in a real chemistry course a strength by enhancing its external validity, this choice could reduce the internal validity of our experiment. Students received instruction on the chemistry content outside our intervention—even though this instruction was the same for all students, it is possible that it interacted with our intervention in an unknown way. Therefore, future research should investigate the effects of sequencing sense-activities and perceptual-activities in more controlled contexts. Third, while we had sufficient statistical power to detect small effect sizes of main effects, our sample may have been too small to detect small interaction effects between sequence and prior knowledge. Therefore, future research should use a larger sample to investigate potentially undetected effects. Fourth, our study did not include control conditions that received only sense-activities, only perceptual-activities, or only regular activities without support for representational competencies. While our prior studies showed that students benefit from the combination of sense-activities and perceptual-activities, those prior studies included short interventions and did not consider progressions across longer learning interventions. Therefore, future studies should test whether the combination of sense-activities and perceptual-activities is indeed most effective throughout the learning progression. In particular, it would be interesting whether in the first stage of our tentative learning progression where we did not find a difference between conditions, students might benefit from regular activities alone, so that they can familiarize themselves with one visual at a time. Finally, future work should investigate how to translate our findings into sequences of sense-activities and perceptual-activities. Specifically, we need to know when a student has reached a level of proficiency that requires a switch in the sequence of these activities. We plan to address this question in our own work by examining how real-time data on students’ problem-solving performance can be used to predict which particular sequence an individual student would most benefit from.

Conclusion
To conclude, our research is, to the best of our knowledge, the first to demonstrate that there may indeed be a learning progression of explicit and implicit representational competencies. Our findings demonstrate that providing activities that support these competencies in a sequence that matches this learning progression has a significant impact on students’ learning of content knowledge. This finding is striking because all students received the same instruction; only the sequence differed. Even when embedded in an existing chemistry course that provided instruction beyond our intervention, the sequence of sense-activities and perceptual-activities affected students’ learning of content knowledge with medium to large effect sizes.
References


Acknowledgments

We thank Judith Burstyn, John Moore, Rachel Bain, and Amelia Berendt for their help with this study. This work was supported by NSF IUSE #1611782.
A Collaboration Script for Nonverbal Communication Enhances Perceptual Fluency With Visual Representations

Martina A. Rau, University of Wisconsin – Madison, marau@wisc.edu
Purav Patel, University of Wisconsin – Madison, ppatel47@wisc.edu

Abstract: Visuals can help students learn only if they understand what visuals show. This involves perceptual fluency: the ability to automatically see meaning in visuals. To support perceptual fluency, instructional interventions support inductive learning by providing simple tasks that require quick attention to relevant features. Sociocultural research shows that collaboration affords nonverbal communication behaviors that helps students attend to visual features more quickly, which helps them acquire perceptual fluency. In contrast, cognitive research suggests that collaboration is ineffective for learning from simple tasks. Hence, we ask: Can collaboration via nonverbal communication enhance the acquisition of perceptual fluency? We investigated this question by examining how 10 dyads collaborate on simple perceptual tasks via nonverbal communication. We analyzed gestures and interviews about learning experiences. Further, we compared dyads’ performance on perceptual activities to 28 students who worked on the activities individually. Our findings suggest that collaboration helps students acquire perceptual fluency with visuals.

Keywords: Perceptual learning, nonverbal communication, collaboration, implicit learning processes, visual representations

Introduction
Instruction in science, technology, engineering, and math (STEM) often uses visual representations (NRC, 2006) such as those in Figure 1. Visuals can help students learn content knowledge by making abstract concepts accessible (Ainsworth, 2006; Rau, 2017c). Yet, visuals can impede learning if students lack representational competencies—knowledge about how visuals show information relevant to scientific and professional practices (Ainsworth, 2006; NRC, 2006). Most instructional interventions focus on sense-making competencies, which allow students to explain how visuals show concepts. For example, chemistry students need to explain how two-dimensional Lewis structures depict a molecule’s three-dimensional geometry. Students acquire these competencies via explicit processes involved in verbal explanation. Recent research suggests that students need a second type of representational competency: perceptual fluency (Kellman & Massey, 2013). Perceptual fluency allows students to quickly and effortlessly infer more than is explicitly shown in a visual. For example, chemists can mentally visualize a molecule’s geometry based on its Lewis structure. Students acquire perceptual fluency via implicit processes involved in inductive and nonverbal learning from experience with many example visual representations (Kellman & Massey, 2013; Koedinger, Corbett, & Perfetti, 2012; Rau, 2017c). Activities that support perceptual fluency have been shown to enhance learning of content knowledge (Rau, 2017c).

Figure 1. Ball-and-stick model (left) and wedge-dash Lewis structure (right) of a molecule.

Prior research has established that collaborative activities afford explicit processes that may enhance students’ acquisition sense-making competencies (Dillenbourg, Baker, Blaye, & O’Malley, 1996; Rau, Bowman, & Moore, 2017). With respect to perceptual fluency, however, extant research provides conflicting views on whether collaboration may enhance or impede students’ learning. In the following section, we review sociocultural research suggesting that collaboration may enhance perceptual fluency and cognitive research suggesting that it may impede perceptual fluency. This review leads to the question: Can collaboration enhance the acquisition of perceptual fluency?
We addressed this question in an exploratory study in which 10 dyads worked collaboratively on instructional activities that support perceptual fluency. These activities engage students in inductive nonverbal learning processes. To investigate whether students’ collaborative interactions align with the learning goals of perceptual activities, we analyzed students’ collaborative interactions. To investigate how students experience collaborating via nonverbal means of communication, we conducted semi-structured interviews. To investigate whether collaborative perceptual activities may be more effective than individual perceptual activities, we used log data generated by the activities to compare students’ performance on perceptual activities to that of students who worked individually on perceptual activities in a prior study.

Prior research

Perceptual fluency with visual representations

Most research on learning with visuals has focused on explicit sense-making competencies (e.g., Rau, 2017c; Stieff, Hegarty, & Deslongchamps, 2011). Yet, recent research suggests that students also need implicit knowledge about visuals, referred to as perceptual fluency (Kellman & Massey, 2013; Rau, 2017c). This argument is based on observations that experts can quickly and effortlessly translate across visuals (Gibson, 2000). Experts “see at a glance” what visuals show without perceived mental effort (Kellman & Massey, 2013; NRC, 2006). They are so efficient at extracting information from visuals that it seems intuitive. For example, chemists automatically see that the visuals in Figure 1 show the same molecule. Perceptual fluency enables experts to effortlessly combine information from multiple visuals, to automatically translate among them (Kellman & Massey, 2013), and to use visuals to communicate with members of scientific and professional communities as if they shared a “visual language” (Kozma, Chin, Russell, & Marx, 2000). Perceptual fluency also allows experts to infer community-specific knowledge beyond what visuals explicitly show (Airey & Linder, 2009).

The sociocultural literature describes perceptual fluency as the product of implicit learning processes that allow students to induce a “visual language” by participating in disciplinary discourse (Wertsch & Kazak, 2011). For example, disciplinary discourse in chemistry often involves translating among multiple visuals (Kozma & Russell, 2005). When participating in such discourse practices, students rarely receive instruction on how to explain translations among visuals but rather infer such translations by imitating how other members of the community use visuals (Kozma & Russell, 2005), often without explicitly knowing what they show (Airey & Linder, 2009; Wertsch & Kazak, 2011). According to the cognitive psychology literature, students acquire perceptual fluency via perceptual-induction processes; that is, they induce connections among visuals without explicit instruction but through experience with many examples (Gibson, 2000; Kellman & Massey, 2013). Perceptual-induction processes are considered nonverbal and implicit because verbal reasoning is not necessary; they can even interfere with the acquisition of perceptual fluency (Schooler, Fiore, & Brandimonte, 1997).

Prior research yields a number of design principles for instructional activities that help students perceptually induce meaning from visuals. Such perceptual activities provide students with many examples in simple tasks that ask students to quickly judge what a visual shows (Kellman & Massey, 2013). Perceptual activities draw attention to relevant visual features by varying irrelevant visual features and contrasting them to relevant features. They encourage students to rely on perceptual intuitions about what they see rather than trying to explain what they see. Students get immediate feedback only on the accuracy of their response without conceptual explanations, so as to engage implicit rather than explicit processes.

Collaborative discourse and perceptual fluency

If students acquire perceptual fluency by participating in disciplinary discourse, it would seem natural to assume that collaborative activities would enhance students’ acquisition of perceptual fluency. Indeed, students often use visuals in collaborative activities (Roschelle, 1992; Wertsch & Kazak, 2011). To date, most research on learning with visuals has focused on explicit sense-making competencies by investigating how students use visuals to collaboratively construct meaningful explanations of concepts (e.g., White & Pea, 2011). By contrast, research on perceptual activities has mostly focused on individual students (e.g., Kellman & Massey, 2013; Rau, 2017a). In the absence of empirical research on how collaborative discourse affects students’ acquisition of perceptual fluency, it is perhaps not surprising that the literature on collaborative learning yields conflicting hypotheses as to whether collaboration may enhance or impede students’ learning from perceptual activities.

On the one hand, sociocultural research on collaborative discourse suggests that collaboration affords implicit processes that enhance perceptual fluency. When collaborating, students use nonverbal means of communication like gesture to direct their partner’s attention to relevant visual features (Singer, 2017). Indeed, analyses of collaborative discourse suggest that pointing at visual features helps students induce meaning of visuals (Rau, 2017b; Stevens & Hall, 1998). When participating in collaborative discourse, students imitate how their
peers use visuals, which helps them become perceptually fluent in disciplinary practices of using visuals for problem solving and communication (Airey & Linder, 2009; Wertsch & Kazak, 2011). Thus, this literature suggests that collaboration could enhance students’ acquisition of perceptual fluency via nonverbal communication.

On the other hand, cognitive research on socio-constructive processes suggests that collaboration may interfere with students’ acquisition of perceptual fluency, for at least two reasons. First, the benefits of collaboration have largely been attributed to verbally mediated co-construction of knowledge (Dillenbourg et al., 1996). Yet, verbalization has been shown to interfere with the acquisition of perceptual fluency (Schooler et al., 1997). Second, perceptual activities involve simple tasks. Research suggests that collaboration enhances learning from complex but not from simple tasks because the latter are most efficiently done individually (Koedinger et al., 2012). Indeed, some studies shows that activities with simple tasks yield lower learning gains if they are done collaboratively than if they are done individually (Kirschner, Paas, & Kirschner, 2010). We note two limitations in these cognitive studies. First, these studies focused on learning explicit knowledge that was acquired via verbal processes. They did not test if collaboration enhances implicit learning of perceptual knowledge that is acquired via nonverbal processes. Second, they did not encourage nonverbal communication. Cognitive studies of individual learning show that nonverbal guidance of visual attention enhances learning with visuals (Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013), but the effects of nonverbal collaboration have not yet been tested.

Research questions
Given the conflicting views of prior sociocultural and cognitive research, our goal is to investigate whether collaboration can enhance perceptual fluency via nonverbal communication. To this end, we conducted an exploratory study that observed students collaborating on perceptual activities while being asked to use only nonverbal means of communication, such as gesture. Specifically, we investigate the following research questions:

1. How do students use gestures to direct each other’s attention to features of visual representations?
2. How do students experience collaborating via nonverbal means of communication?
3. How does the acquisition of perceptual fluency compare between students working collaboratively versus individually on perceptual activities?

Method
Participants
Twenty bachelor and master students from a large university in the Midwestern U.S. participated in this study. They were recruited from undergraduate chemistry courses and flyers distributed across campus. Due to the exploratory nature of the study, their experience with chemistry ranged from no formal instruction to advanced undergraduate courses. Dyads were formed based on students’ scheduling constraints.

Perceptual activities

Students worked on perceptual activities using an educational technology for undergraduate chemistry—Chem Tutor. These activities have been shown to enhance perceptual fluency (Rau, 2017a). The goal of these activities is to engage students in inductive processes. Figure 2 shows an example perceptual activity. Students are given one visual (Figure 2, left-hand side) and have to select one of four visuals that shows the same molecule (Figure 2, right-hand side). Students are prompted to use gesture to agree on an answer. If they make an error, the choice is highlighted in red and students receive a prompt to agree on a new answer by gesture. Students point at visual features such as atoms and bonds.

Figure 2. Example of students working in dyads on a collaborative perceptual activity.
analyses to compare error rates and answer-duration measures to those obtained from 28 undergraduate chemistry students who worked on the activities as an individual homework assignment for a chemistry course. Van Gog and Paas (2008) as: efficiency = (z-score of accuracy – z-score of duration) / √2. We used statistical combined these metrics in an efficiency measure of how long it took students to achieve a correct answer, following (the start of the activity or an incorrect answer in the same activity until the current answer). Finally, we com-
considered important. Specifically, the goals of the perceptual activities are to direct students' visual attention to 2), we qualitatively examined interview responses for themes that emerged across dyads.

Nonverbal visual attention direction
We investigated how students use nonverbal attention direction to address research question 1. We used grounded analysis to identify codes to describe students’ gestures. Three visual features emerged. Gestures were coded as number of atoms if students pointed at multiple atoms in a way that indicated counting, for example by indicating a number with their other hand. We coded gestures as distinct atoms if students pointed at particular atoms or groups of atoms that differed from other atoms in a molecule (e.g., a chlorine atom when all other atoms were carbons or hydrogens). Gestures were coded as bonds if students pointed at lines connecting atoms. In addition, we coded a gesture as answer if students pointed at an entire image rather than a particular feature. Further, we observed that on later activities, students seemed to point more often only at the entire answer im-

Procedure
Before students worked on the perceptual activities, they received a brief introduction explaining what percept-
tual learning is. Next, they were asked to try out this procedure on an unrelated perceptual activity that used different visuals. They were instructed to collaborate on these activities using only nonverbal communication such as gestures. They were told that they were not allowed to talk during the activity. Once comfortable with the method and after being given the opportunity to ask questions, students received 20 consecutive perceptual activities. The collaborative session was followed by a semi-structured interview during which students were asked to reflect on their experience with nonverbal communication during the collaborative session. The interviewer asked predefined questions about what was difficult, what was helpful, and whether gesturing helped them learn. The interviewer asked both participants, alternating which of the two students was to answer first. The interviewer also asked follow-up questions to clarify or expand on students’ responses. All sessions were videotaped and interviews were transcribed.

Measures and analyses
To investigate how students used gestures to direct each other’s attention to features of visual representations (research question 1), we segmented videos by perceptual activity. For each consecutive gesture (i.e., from the initiation of the hand movement to its finish), we coded which visual features students pointed at. To develop the coding scheme, we used a grounded approach to identify features that emerged across dyads. Interrater reliability, determined based on 10% of the data, was high (Kappa = .96). We compared these features to the learning goals of the perceptual activities, as determined by our prior research on novice students and expert chem-
ists. The learning goals correspond to visual features that chemistry students often failed to attend to but experts considered important. Specifically, the goals of the perceptual activities are to direct students’ visual attention to (1) distinct atoms in a molecule, (2) the total number of atoms in a molecule, and (3) the bonds between atoms.

To investigate how students experience collaborating via nonverbal communication (research question 2), we qualitatively examined interview responses for themes that emerged across dyads.

To investigate how collaborative acquisition of perceptual fluency compares to individual acquisition of perceptual fluency (research question 3), we computed error rates as the number of incorrect first attempts for each activity. Further, we computed answer-duration as the time between the previous and the current action (the start of the activity or an incorrect answer in the same activity until the current answer). Finally, we com-
bined these metrics in an efficiency measure of how long it took students to achieve a correct answer, following Van Gog and Paas (2008) as: efficiency = (z-score of accuracy – z-score of duration) / √2. We used statistical analyses to compare error rates and answer-duration measures to those obtained from 28 undergraduate chemistry students who worked on the activities as an individual homework assignment for a chemistry course.

Results
Due to a technical issue, one of the ten dyads was missing video data for the first three of the 20 activities.
age. Therefore, we also recorded whether an activity had only answer codes. Finally, we observed that students achieved an answer with fewer gestures over time. Therefore, we analyzed how the proportion of gesture types per activity changed over time by dividing the frequencies of codes by the number of total gestures used in the given activity. Table 1 provides an overview of the resulting metrics, categorized by quartiles of progress through the 20 perceptual activities.

Table 1: Visual features students pointed at during by quartile of their progress through the perceptual activities: average number of total gestures per activity; proportion of gestures that pointed at number of atoms, distinct atoms, bonds, or answer, and proportion of activities with answer gestures only

<table>
<thead>
<tr>
<th>Quartile (Activities)</th>
<th>Gestures per activity</th>
<th>Number of atoms</th>
<th>Distinct atoms</th>
<th>Bonds</th>
<th>Answer</th>
<th>Answer-only activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (1-5)</td>
<td>3.11</td>
<td>10.2%</td>
<td>13.6%</td>
<td>16.3%</td>
<td>56.1%</td>
<td>12.3%</td>
</tr>
<tr>
<td>2 (6-10)</td>
<td>2.32</td>
<td>8.6%</td>
<td>11.2%</td>
<td>15.5%</td>
<td>56.9%</td>
<td>13.8%</td>
</tr>
<tr>
<td>3 (11-15)</td>
<td>2.46</td>
<td>11.4%</td>
<td>8.9%</td>
<td>8.1%</td>
<td>52.9%</td>
<td>13.8%</td>
</tr>
<tr>
<td>4 (16-20)</td>
<td>2.18</td>
<td>15.6%</td>
<td>4.6%</td>
<td>12.8%</td>
<td>59.7%</td>
<td>20.2%</td>
</tr>
</tbody>
</table>

Given the small number of gestures per activity, a quantitative analysis of the gestures was not warranted. Instead, we qualitatively inspected patterns and trends over time. Overall, it is striking that students exclusively pointed at visual features that match the learning goals of the perceptual activities. Somewhat to our surprise, students did not point at other features such as bond angles or branching. Further, over time, we observed that students paid more attention to the number of atoms, whereas they paid less attention to distinct atoms and bonds. We also found that over time, dyads required fewer gestures to agree on an answer. Finally, students seemed to converge on more holistic processing over time, pointing more frequently at the entire answer choice and more frequently pointing only at the answer without additionally pointing at other features.

Student experiences with nonverbal communication

To address research question 2, we examined how students experienced nonverbal communication when collaborating on perceptual activities. The following themes emerged. First, with respect to comfort, several students commented that gesturing was uncomfortable at first but got easier with practice: “I feel like at the beginning, just ‘cause we didn’t really know exactly what the other person was saying, it was a little challenging, but towards the end we kind of like figured out, what we meant and then we ended up both doing the same gestures meaning for the same thing.” Others reported being comfortable with gesturing: “I feel like that’s easier than you telling me ‘Oh, look at the green…’ see I don’t have words for the diagram, so, I think it was easier to just like point out why that was.” Several students reported enjoying gesturing: “It felt like, a game almost.”

Second, regarding helpfulness, all students reported that their partner’s gestures helped them attend to relevant visual features: “you wouldn't have gotten it right on the, first try, you were pointing to something and you were like Oh yeah, I shouldn't have missed that.” Students explained why they found gesturing helpful because it matched the perceptual nature of the task: “For this specific task, because it has to do with orientation, like how stuff is positioned, gestures often help versus just saying pointing up.” Also, students found gestures helpful because they made them feel more confident about their perceptions than if they had worked on the activities individually: “I think I would have doubted myself more because I don't know a lot about chemistry.” Further, students thought that gestures aligned with nonverbal communication in chemistry: “I think it reminded me that nonverbal communication is very key.”

Third, students had conflicting experiences with respect to the efficiency of the collaboration process. Some students found collaboration inefficient. For example, when asked how the task would have been different if they had worked on the same activities alone, one students said: “I think I would have been faster but probably less accurate.” By contrast, her partner expected to be more efficient when collaborating: “I think for me it’s the opposite I think I would have taken my time and kind of gone through it more slower.” Many students mentioned that they became more efficient at gesturing as a result of learning what features to attend to: “then he didn't have to point to it later on for me to be like the green is Cl.”

Acquisition of perceptual fluency from collaborative versus individual activities

To address research question 3, we examined how students’ acquisition of perceptual fluency differed between students working collaboratively versus individually. A t-test with error rates as dependent measure and collaborative versus individual work as independent factor showed a large effect such that students who worked collaboratively had significantly lower error rates, \( t(36) = 8.27, p < .01, d = 1.05 \). Further, we inspected students’ learning curves, that is, how their error rates decreased as they progressed through the perceptual activities. Fig-
Figure 3 shows that students’ error rates decreased more quickly as a function of their progress through the perceptual activities when they worked collaboratively (purple) than when they worked individually (orange). Next, a t-test with answer-duration as dependent measure showed a large effect of collaborative work, such that students took significantly more time to submit an answer when working collaboratively, \( t(36) = 5.92, p < .01, d = 3.18 \). Finally, a t-test with efficiency as dependent measure showed no significant effect of collaborative work \( (t < 1) \).

**Discussion**

On the one hand, sociocultural research on collaborative discourse suggests that nonverbal communication is a key process through which students acquire perceptual fluency in collaborative settings (e.g., Wertsch & Kazak, 2011). On the other hand, cognitive research on collaborative learning suggests that collaboration may interfere with students’ learning. Specifically, this research has argued that (1) collaboration can reduce the effectiveness of perceptual learning because communication interferes with inductive processes (Schooler et al., 1997) and (2) collaboration can reduce the efficiency of perceptual learning because simple tasks are more efficiently done individually (e.g., Kirschner et al., 2010; Koedinger et al., 2012).

Our findings suggest that collaboration may indeed enhance students’ acquisition of perceptual fluency from activities that support perceptual fluency with simple tasks. First, we found that students can use nonverbal communication to direct each other’s attention to visual features that align with the learning goals of perceptual activities by pointing at the number of atoms in a molecule, distinct atoms, and bonds. Over time, students increasingly prioritize the number of atoms. Further, their gestures become more efficient: it takes students fewer gestures to achieve the right answer. Finally, students’ gestures become more holistic over time: students point at the entire answer picture rather than at specific features. Because we had identified the learning goals of the perceptual activities based on features that expert chemists consider to be important, these findings suggest that students’ gestures align with expert discourse practices.

Second, we found that after becoming familiar with the method, students felt comfortable with gesturing. Further, they experienced their partner’s gestures as helpful for four reasons: (1) they helped them attend to visual features they otherwise would have missed, (2) they increased their confidence in their perceptions, (3) they align with the perceptual nature of the activities, and (4) they aligned with the use of nonverbal communication in chemistry discourse. With respect to efficiency, students’ experiences were mixed: while some commented that they found nonverbal communication inefficient and would have preferred to talk, others thought it helped them become more efficient at the perceptual activities than they would have been when working alone.

Third, our comparison of the log data from this study to a prior individual study corroborates the analysis of the gestures and interviews. Specifically, we found that students indeed achieved higher accuracy in translating among the visuals when working collaboratively than when working individually. However, it took students longer to arrive at their answers when they worked collaboratively. When we combined accuracy and duration metrics into an efficiency measure, we found no evidence that the longer response times made students less efficient in collaboratively achieving a correct answer, compared to students working individually.

Our findings are surprising in light of prior cognitive psychology research. This prior research suggested that collaboration may reduce the effectiveness of perceptual activities because communication interferes with perceptual learning processes (Schooler et al., 1997). However, this prior research focused on verbal communication. Our findings suggest that nonverbal communication may actually enhance (rather than interfere) with perceptual learning. Further, prior research claims that collaboration may reduce the efficiency of learning...
from perceptual activities that provide students with many examples in simple tasks. Indeed, Koedinger and colleagues (2012) describe collaboration as an instructional intervention that enhances sense-making processes, which are time-consuming and therefore reduce the efficiency of students’ learning from simple tasks that do not require sense making. Our findings agree that collaboration takes time, but—counter to the argument by Koedinger and colleagues—we found no evidence that the increase in duration came at the expense of overall efficiency. Rather, increase in duration is associated with higher accuracy. These quantitative findings corroborate the qualitative analyses of gestures and interviews that suggest that the extra time required to collaborate is worthwhile because it is associated with higher confidence and because students learn to engage in nonverbal communication practices that align with what expert chemists pay attention to when working with visuals.

To summarize, our study makes two important theoretical contributions to prior research on collaborative learning. First, we provide empirical evidence for the argument in the sociocultural literature that nonverbal communication is a key process through which students collaboratively induce meaning of visuals. We readily concede that one can hardly consider our study setup a representative example of disciplinary discourse—being a rather artificial learning situation in a research lab. Yet, we believe that our ability to show that nonverbal communication can enhance perceptual fluency even in an artificial setting all the more underscores the importance of collaboration for perceptual learning. Second, we expand prior cognitive research on collaborative learning by showing that simple tasks focused on implicit, nonverbal knowledge can be learned more effectively in collaboration with partners, provided that the activities focus on implicit, nonverbal knowledge and encourage nonverbal rather than verbal communication. This finding suggests that there may be a new boundary condition for the effectiveness of collaborative learning from simple tasks.

Our study makes two practical contributions. First, it suggests that to support students’ acquisition of perceptual fluency, instruction should not disregard collaborative activities. We find that collaborative perceptual activities can enhance students’ acquisition of perceptual fluency. In light of prior cognitive research, however, we recommend encouraging students to communicate via nonverbal means like gesture. Second, our findings suggest that collaboration scripts that prompt students to use gestures to collaborate and prohibit verbal communication may be effective at enhancing collaborative learning from simple tasks—as opposed to complex tasks that benefit from collaboration scripts that support verbal communication.

Limitations and future directions
Our contributions should be interpreted against the following limitations. First, our study was conducted in a research lab. This is not representative of most realistic collaborative learning activities. Especially in light of the argument that perceptual learning occurs while students engage in disciplinary discourse practices, future research should investigate whether prompting students to use nonverbal communication enhances their acquisition of perceptual fluency in realistic learning contexts. Second, we did not assess students’ prior knowledge or their prior perceptual fluency. Future research should include additional assessments to investigate whether collaboration may differentially benefit students with low or high prior knowledge. Third, we did not investigate whether collaborative perceptual learning helps students participate in verbal disciplinary discourse. Future studies could investigate this question by observing how students communicate verbally after having worked on perceptual activities either individually or collaboratively. Fourth, we did not test whether collaborative perceptual activities are more effective than individual perceptual activities at enhancing learning of content knowledge. Given that perceptual fluency enhances learning of domain knowledge (e.g., Kellman & Massey, 2013), we expect that the enhanced acquisition of perceptual fluency that we observed when students were learning collaboratively increases their acquisition of content knowledge. Future research should investigate this hypothesis by assessing students’ individual content knowledge. It would be interesting to test effects on content knowledge immediately after collaborative perceptual activities and after subsequent instruction that uses the same visual representations (e.g., ball-and-stick models and Lewis structures) for different chemistry concepts.

Conclusions
Our exploratory study provides valuable directions for future research. It demonstrates that collaboration via nonverbal communication can enhance perceptual learning in a controlled setting. We believe this finding also demonstrates the importance of nonverbal perceptual processes for collaborative learning. In our opinion, cognitive psychology research was too quick to dismiss the potential benefits of collaborative activities for learning from simple tasks because it did not consider nonverbal perceptual learning. By addressing this gap in the literature, our study bridges sociocultural and cognitive research on collaborative perceptual learning. In general, our results illustrate that cognitive psychology research on collaborative learning can benefit from aligning instructional supports with disciplinary practices documented in the sociocultural literature.
References

Acknowledgments
We thank research participants and research assistants. This work was supported by NSF IIS 1623605.
Teachers’ Mediation of Students’ Interactions with Physical and Virtual Scientific Models in Biology

Nicole D. Martin, Dana Gnesdilow, and Sadhana Puntambekar
ndmartin@wisc.edu, gnesdilow@wisc.edu, puntambekar@education.wisc.edu
University of Wisconsin – Madison

Abstract: This study explored the mediating role that teachers play in helping students utilize multiple models as tools for scientific thinking. We analyzed two teachers’ discourse as they facilitated students’ use of virtual and physical models and compared their students' understanding of and engagement with models during a biology unit. We found that one teacher had significantly more talk than the other about the affordances and constraints of models, and his students more purposefully used the affordances of the models to engage in scientific thinking on their own and seemed to learn more about affordances and constraints of scientific models than students in the other teacher’s class. Our findings suggest that the ways in which teachers talk about the affordances and constraints of scientific models is important for students to understand and utilize them as tools for thinking and building knowledge.

Introduction
Models and modeling play a central role in the practice of science, and teaching students to understand and use scientific models is an important goal of science education (Lehrer & Schauble, 2006; National Research Council, 2012; Windschitl, Thompson, & Braaten, 2008). Scientists use models to represent particular aspects of phenomena and to generate, test, and revise scientific ideas. In this way, scientists employ models as tools for thinking and building knowledge. Scientific models such as physical representations and lab experiments have been pervasive in classrooms, and other types of models such as virtual simulations are becoming increasingly prevalent. However, modeling is a complex practice, and the widespread presence of models in science classrooms does not mean that students know how to use them as tools for scientific thinking. Research has suggested that students often do not understand the practice of modeling in science, and teachers need to support students in learning this practice (Schwarz et al., 2009; Treagust, Chittleborough, & Mamiala, 2002).

From a sociocultural perspective, learning is mediated by interactions with cognitive tools and cultural resources (Vygotsky 1978, Wertsch 1991). Scientific models are cognitive tools that mediate thinking and have particular norms and practices that guide their use. Tools contain the intelligence of the maker (Pea, 1993), and by using a tool, it is possible to capitalize on and benefit from the distributed intelligence embedded in the use of the tool. Thus, tools can support thinking and reasoning in ways that may not be possible without the aid of the tool (Pea, 1993). Additionally, tools are designed to have particular affordances and constraints that influence their use. While the affordances of a tool are features intended to draw the user’s attention and guide effective use, such affordances are not always obvious to the user (Norman, 1988). Thus, the tool may not be used in ways that fully benefit the user. Further, being able to use or learn from a tool also requires understanding the cultural norms around its use (Cole & Engeström, 1993). From this perspective, using scientific models not only requires understanding the affordances of particular models, but also how the scientific community uses such tools to generate, test, and revise scientific ideas to build knowledge.

Teachers can play an important role in mediating students’ engagement with models to successfully utilize their affordances and realize their potential for supporting students’ learning. As discussed by Cole and Engeström (1993), while tools can mediate students' thinking, the ways in which students interact with and use a tool can be further mediated by a teacher. This mediation is often manifested in teachers’ discourse as they interact with and support students during activity in the classroom. However, prior work suggests that teachers may not effectively mediate students’ interactions with tools in classrooms (Kozulin, 2003). This may be due to teachers’ beliefs or assumptions that the meaning and affordances embedded in tools are explicit enough for students to understand on their own, and thus do not necessitate mediation (Kozulin, 2003), or due to teachers themselves not understanding the meaning and affordances of tools.

Additionally, mediation from a teacher may be increasingly important in environments where different types of models or tools are being employed to help students learn. In science, different types of scientific models can provide unique affordances and are often used to provide students with alternative ways to experience scientific phenomena. For example, virtual models such as simulations afford altering the time and visual scale of phenomena to allow for fast experimentation about unobservable or slow processes and allow students to constrain complex systems (de Jong, Linn, & Zacharia, 2013; Olympiou & Zacharia, 2012). Physical
representations, on the other hand, can expose students to authentic problems and complex systems about real phenomena, provide concrete qualitative and quantitative observations or measurements, and raise challenges students would not face in a virtual experiment (de Jong et al., 2013; Olympiou & Zacharia, 2012). Previous research on using models in science has explored students’ learning from conducting experiments with either virtual or physical models, as well as participating in different sequences of virtual and physical experiments (e.g., Zacharia & Olympiou, 2011). This work has yielded mixed results about the most beneficial modality and sequencing for learning science content. Rather than focusing on one type of model, using virtual and physical models together offers the potential of capitalizing on the unique affordances of both types of models to support students’ learning (Olympiou & Zacharia, 2012). Initial investigations into how students are able to coordinate information from virtual and physical models suggest that understanding and utilizing the affordances of these models may be quite challenging for students (Martin, Gnesdilow, & Puntambekar, 2017). Thus, while there is great potential for utilizing the unique affordances of both virtual and physical models to support students’ learning, these previous findings suggest that this potential will not be reached without proper support and orchestration by the teacher. Despite calls for teaching students via modeling in K-12 science education and research showing that students have difficulty using models, little is known about how teachers actually mediate students’ effective use of models in the science classroom.

Given the difficulties students’ face in understanding and coordinating virtual and physical models to take advantage of their unique affordances, we explored the mediating role that two teachers played in helping students to utilize multiple models to learn science. Specifically, we aimed to answer the following two-part question: How do teachers mediate students’ interactions with multiple models to help them: a) use models as tools for scientific thinking and, b) understand the affordances of different models to learn science? To explore this question, we analyzed two teachers’ discourse as they facilitated students’ use of virtual and physical models and compared their students’ understanding of and engagement with models during a biology unit.

Methods

Participants and context

The participants in this study were two experienced 8th grade science teachers, Mr. Fox and Mr. Smith, and the students in each of their science classes (n = 27, n = 21 respectively). The teachers taught at different urban middle schools in the same U.S. Midwestern city. The students at Mr. Fox’s school were 64% white, 18% Hispanic, and 8% black or African American, and 55% were economically disadvantaged. The students at Mr. Smith’s school were 78% white, 10% Hispanic, and 4% black or African American, and 40% were economically disadvantaged. Mr. Fox’s class happened to be designated as a “challenge” class, in which students were often given more open-ended tasks. Mr. Smith’s school did not offer challenge classes, thus his class consisted of students with a range of abilities including students who could have succeeded in a challenge class. This was a design-based research study in which we did not purposefully select these two classes for comparison a priori, but rather we selected these teachers because we had a complete corpus of teacher and students data to pursue our research question.

This study was a part of a larger project aimed at helping students learn biology by engaging them in solving 21st century real-world bio-engineering problems. Both teachers taught the same 10-12 week design-based, biology curriculum that incorporated the use of both virtual and physical models. The curriculum challenged students to design a compost that would break down quickly and contain a lot of nutrients, to reduce the amount of waste going into landfills. Students learned key concepts related to energy transformation and matter cycling in ecosystems to solve their challenge. To study decomposition and collect data to justify their compost designs, students worked in small groups of three to four to build, monitor, and refine a physical bio-reactor during the unit. Given that decomposition takes several weeks, towards the middle of the unit students also used a virtual compost simulation to investigate how abiotic factors influence decomposers’ ability to break down matter. In their small groups, students conducted three sequential virtual compost experiments related to the carbon to nitrogen ratio, amount of moisture, and particle size of the materials in compost. Based on their findings from the virtual experiments, students were asked to revisit their physical bio-reactors and use the science ideas and data from the virtual experiments to make decisions and justify a change to increase decomposition in their physical bio-reactors. Teachers facilitated small-group and whole-class discussions as students engaged in these activities throughout the unit.

Data sources and analysis

We used three data sources: i) classroom videos of teachers’ interactions with their students as they conducted all three virtual experiments, ii) audio-recorded student conversations during the last virtual experiment, and iii)
students’ pre and post scores on an open-ended question about the affordances and constraints of physical and virtual models in science, taken at the start and end of the unit.

We examined the classroom videos of both teachers to investigate each teacher’s discourse as they mediated students’ work with the virtual compost simulations and physical compost bio-reactors. We selected particular lessons in which students used the virtual simulations to investigate factors affecting decomposition in their physical bio-reactors, because these lessons captured opportunities for students to interact with both their physical and virtual compost models and allowed us to investigate how teachers mediated students’ thinking and use of multiple models to learn science. We used a mixed-methods approach to quantify the qualitative discourse data present in the videos (Chi, 1997). The videos were transcribed and teachers’ turns of talk were deductively and inductively coded to capture how teachers supported students’ interactions with the virtual and physical models. Based on previous literature suggesting that students likely need support to learn about the nature of models (Schwarz et al., 2009), the affordances and constraints of models (Kouzlin, 2003; Norman, 1988), and how to coordinate multiple scientific models (Martin et al., 2017), we were particularly interested in how teachers discussed the general nature of models, the affordances and constraints of virtual and physical models in this context, and connections between the models. Additionally, we were interested in how teachers used the virtual and physical models to support students’ learning of science content and practices. This resulted in the five mediation codes described in Table 1. The nature of models code was adapted from our prior work on teachers’ discourse related to using models (Gnesdilow, Smith, & Puntambekar, 2010), while the codes for affordance and/or constraint of virtual or physical models, connecting models, science content, and science practices were developed to specifically capture important discourse about the virtual and physical models in the context of this study. Some talk was coded as not applicable, which often consisted of managing procedural and social interactions. Turns of talk could receive multiple codes if multiple forms of mediation discourse were present. Ten percent of the teachers’ turns of talk were coded by the first and second author and an 84% agreement was achieved. The remaining discourse was coded by the second author. We then calculated the proportion of the different types of talk that both teachers engaged in as they mediated students’ use of the models. These proportions accounted for the different amounts of overall teacher talk during all the virtual experiments, so the proportion of each type of talk could be compared between teachers. To do this, we divided the frequencies for each type of mediation discourse from Mr. Fox and Mr. Smith by the total number of turns of talk for that teacher. We then conducted a Chi-squared test of homogeneity of proportions to compare the prevalence of talk related to mediating students’ use of the virtual and physical models between teachers.

Table 1: Coding scheme for teachers’ discourse to mediate students’ interactions with models

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of Models</td>
<td>Talk referring to scientists’ construction and use of models, roles of models in science, models as abstractions, model evaluation</td>
<td>“What’s a model? Your bio-reactors right? Alright, and your bio-reactors are a model for what? What are they trying to model for you, what are they trying to show you?”</td>
</tr>
<tr>
<td>Affordance and/or Constraints of Virtual or Physical Models</td>
<td>Explicit or implicit talk about usefulness of virtual model to: quickly run multiple trials; constrain the complex system by isolating variable; input current physical conditions into the simulation to quickly run trials to make predictions</td>
<td>“…What’s the advantage of using a computer simulation of a compost pile, or of a compost? …We can control it better because we don’t have variables outside of what we’re just looking at. Ok? So, that’s what these, uh, virtual experiments are meant to do.”</td>
</tr>
<tr>
<td></td>
<td>Explicit or implicit talk about usefulness of physical model to: understand, explore messiness of complex system in real world; raise challenges; provide concrete observations; quantitative &amp; qualitative measurements</td>
<td>“You gotta think about this and doing it on a large scale, too. Remember the challenge. The challenge is: compost, doing this at our school, with, fast, and a lot of nutrients. Right? Well, what would happen to a compost pile at school? With regards to, adding greens or browns?”</td>
</tr>
<tr>
<td>Connecting Models</td>
<td>Explicit or implicit talk about how to use both virtual and physical models in tandem to synthesize ideas and learn science. Helping students to see connections between information provided by both models, or to use information provided by one model to guide or justify usage of other model.</td>
<td>“After you run your particle size (virtual experiment) then you as a group need to talk about what changes you think you could make or what problems you have (in bio-reactor) …What problems do you have, what could you do about it…what’s your evidence, and then what one do you think is going to improve”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“And then we can use that information, that you find in the (virtual) experiment, and apply it to what</td>
</tr>
</tbody>
</table>

© ISLS
you’ve got going on in your bottles. Ok?”

<table>
<thead>
<tr>
<th>Science Content</th>
<th>Using results or information from one or both models to help students understand science concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Practices</td>
<td>Using one or both models to scaffold students’ ability to conduct fair tests, collect accurate data, analyze, and draw conclusions</td>
</tr>
</tbody>
</table>

“What do the bacteria in your decompos- the bacteria that are acting as decomposers, what are they doing inside your compost? What are they doing with the carbon?”

“You’re gonna start looking at these variables that we can control and, and, manipulate. If we’re controlling one, say the carbon to nitrogen ratio. Should we, change, where we’re placing the bioreactor?”

To investigate how students’ discussions around the virtual and physical models may have related to the mediation provided by the teachers, we also analyzed conversations from all the small groups of students that we had audio-recorded during the unit (n=3 for Mr. Fox, and n=3 for Mr. Smith). These groups were chosen for audio-recording by each teacher at the beginning of the unit as representative, average-performing groups in the class. We selected students’ conversations from the third virtual experiment, in which students investigated how the size of particles in compost affected the rate of decomposition. We chose this last experiment as an outcome measure to examine how students used the virtual and physical models as tools for their inquiry, since they had received mediation from their teacher during the prior two experiments. When examining students’ discourse during the experiment, we looked to identify any patterns or themes in the types of interactions that students engaged in as they used the models to better understand if there were any key differences in the ways that the students in each teacher’s class used the virtual and physical models as tools for scientific thinking.

Additionally, we used students’ pre and post scores on an open-ended question about the affordances and constraints of physical and virtual models in science. The question presented students with a scenario in which one student conducted an experiment using a virtual computer simulation and another student conducted the same experiment using physical models. Students were then asked to describe reasons why the two students in the scenario might have gotten different results from their experiments. Students received a score of 0 for responses that were incorrect or lacked any mention of an affordance of either type of model, a score of 1 for responses that discussed one affordance that could contribute to different results, and a score of 2 for responses that discussed two or more affordances. We used a non-parametric analysis of covariance (Young & Bowman, 1995) to compare students’ post-test scores based on the teacher they had, with students’ pre-test scores as the covariate. This test was chosen because our dependent variable (post score) was categorical and our sample size was small, and thus the data were non-normal.

Results

Teachers’ mediation of students’ interactions with models

To better understand how the teachers mediated students’ interactions with multiple models to promote the utilization of the affordances of different models to learn science, we conducted a Chi-squared test of homogeneity of proportions to compare differences in teachers’ talk within our 5 coding categories.

![Figure 1. Proportion of teacher talk to mediate students’ interactions with virtual and physical models.](image)

As seen in Figure 1, our results showed that Mr. Fox had a significantly higher proportion of talk about the **affordances and constraints of models** (0.05) than Mr. Smith (0.03) (p < .05). Additionally, Mr. Smith had a significantly higher proportion of talk about learning **science content** (0.52) than Mr. Fox (0.29) (p < .001).
However, there were no significant differences in the proportions of talk that Mr. Fox and Mr. Smith engaged in about science practices (0.17 & 0.15, respectively), making connections between the models (0.08 & 0.07, respectively), and the nature of models in general (0.005 & 0.004, respectively).

Students’ use and understanding of models

Examination of groups’ discourse

We examined groups’ conversations (n=3 groups for Mr. Fox, n=3 groups for Mr. Smith) to understand how students’ discussions about and interactions with the virtual and physical models may have been influenced by the mediation they received from their teacher. By examining students’ discourse while they conducted their particle size simulation experiment, we identified several interaction patterns that revealed key differences in how students used models in Mr. Fox’s and Mr. Smith’s classroom. A major theme we identified was that Mr. Fox’s students showed several instances of purposefully using the affordances of the models to engage in scientific thinking, while Mr. Smith’s students rarely took advantage of these affordances and, instead, focused on procedural aspects of completing the experiment and getting a “good result.” Below, we provide a general overview of the groups’ discussions in each teacher’s class and an excerpt from one group in each teacher’s class to clearly illustrate the differences in the types of discussions that groups had depending on their teacher.

During the particle-size virtual experiment, all three groups of students we audio-recorded in Mr. Fox’s class had conversations about inputting conditions from their physical model into the virtual simulation to make connections between the two models, and they utilized the affordances of the virtual model to explore science content. For example, in the excerpt below, one group used the visualizations presented in the virtual simulation to test their ideas about how air circulation might affect the rate of decomposition in compost. To do this, they took advantage of the affordance of quickly comparing the visualizations in the model of both smaller and larger particle sizes to analyze, make inferences, and explain how the amount of space between particles might affect air circulation in compost. They then used this information to decide on an optimal particle size of 50mm for their physical bioreactor.

Student 1: Did they talk about the air circulation in this simulation?
Student 2: Um, we talked about it the other day. I don’t think it talked about it in the simulation. But I mean, here, make it really small. Like make the particle size really small. Make it 5. And then put some in. And then make it 100.

Student 1: What?
Student 2: Now make it 100.
Student 1: Right now?
Student 2: Yeah.
Student 1: You can’t.
Student 2: Yeah you can.
Student 1: It just deletes. (Note: the simulation only allowed students to use one size particle per trial, deleting previous particles if a student tried to add multiple sizes to the virtual compost)

Student 2: Oh. Well you saw how much was in there for 5?
Student 1: Yeah.
Student 2: And then, you know, you go to 100…
Student 1: And there’s more pockets, there’s more space.
Student 2: Exactly. So there’d be better air circulation the bigger the particle size.
Student 1: But if it’s too big there’s too much air and not enough moisture.
Student 2: Exactly. So you’d have it, like 50.
Student 1: 50 would be perfect. ‘Cause it’s in between 15 and 75.

This excerpt exemplifies how Mr. Fox’s students used the virtual simulation not only to complete their experiment, but to actually capitalize on the visualizations afforded by the simulation to use this model as a tool for scientific thinking. Rather than just focusing on the size of the particles, students discussed the concept of air circulation and its relationship to the rate of decomposition. Further, they chose to compare the amount of space created between the smallest and largest particle sizes allowed by the simulation, instead of choosing random particle sizes, to optimize the visualization to test their ideas. Understanding the relevance of air circulation in
the simulation was important for being able to think about air as an abiotic factor that influenced decomposition in their physical compost bio-reactor. The students used the simulation as a tool to think through these ideas.

In contrast, the groups we examined in Mr. Smith’s class predominately capitalized on the efficiency of the virtual simulation to quickly conduct trials and find the particle size range that was ideal for decomposition so they could answer the questions at the end of the lab. Overall, the students seemed to use the virtual simulation as a game to find an ideal result, rather than as a tool to help them think more deeply about the science content or test their science ideas. For instance, in the excerpt below from one group, Student 1 described how he was “messing around” with the simulation, happened to get a “good result,” and was excited that he had “won.” When Mr. Smith interacted with this group shortly after, it was clear that the students’ main goal was to find the upper and lower bounds of the ideal particle size range, without using the affordances of the virtual model to think about what was important about particle size from a scientific perspective.

Student 1: Okay. I just gotta, I’m just messing around this first time okay?
Student 2: Whoa.
Student 1: I actually did good. I did good. I did good. I won. I just put it to a random size and shook it all in there and-
Student 2: It was at like 74. Yeah, see 74. That’s our first simulation.
Student 3: Set it there.
Student 1: I didn’t even try, how did I, like what? Particle size 74…Rate of decomposition fast.
Temperature, ideal. Odor, normal…What is the result? It’s a good result.

Mr. Smith: So what are you guys thinking? What have you guys found?
Student 1: Um the smallest is way too slow to decompose.
Mr. Smith: And what was the smallest?
Student 3: 5mm, right? Millimeters that’s the double m?
Mr. Smith: Millimeter. Yeah. So you guys started with, what is that?
Student 1: 74.
Mr. Smith: 74 is where you started?
Student 1: Yeah. So then 50 now it goes back-
Mr. Smith: So you guys are all over. So you guys were just trying to hit the upper and lower immediately?
Student 2: Yeah.
Student 1: Yeah I was messing around the first time and then it actually worked.

This excerpt is one of many that could have been used to demonstrated that students in Mr. Smith’s class commonly used the simulation in a trial-and-error fashion, rather than thoughtfully identifying parameters to test their ideas in a scientific way. Further, the students in this group and the other groups in Mr. Smith’s class never talked about why different particle sizes influenced the rate of decomposition. For example, they never used the simulation to go beyond finding the "right" answer and to discuss the cause-and-effect relationships between particle size and aerobic versus anaerobic conditions or how larger particles have less surface area for decomposers to break down. The simulation gave students comments, questions, and visualizations to help them think more deeply about factors that influence decomposition, but they did not utilize these affordances.

Examination of students’ pre - post performance on open-ended question about affordances and constraints of models
To further examine if there were differences between Mr. Fox's and Mr. Smith's students' understanding of the affordances and constraints of virtual and physical models, we compared students’ pre and post responses to an open-ended question about the affordances and constraints of virtual and physical models. We used a non-parametric analysis of covariance (Young & Bowman, 1995) due to our small sample size. When using the pre scores as a covariate, we found that Mr. Fox's students' post scores were significantly higher than Mr. Smith's students' scores (p < .01). This showed that Mr. Fox's students learned more about the affordances and constraints of virtual and physical models than Mr. Smith’s students.

Discussion
Research suggests that the use of tools can be further mediated by a teacher to support students' learning (Cole & Engeström, 1993). Scientific models are tools that help students engage in the practices of science. Therefore,
the ways in which teachers talk about the affordances and constraints of scientific models is important for students to understand and utilize them as tools for thinking and building knowledge. Our findings supported this idea and provided evidence that the type of mediation students received from their teacher may have influenced their understanding of and engagement with models. Mr. Fox engaged in a greater proportion of talk about the affordances and constraints of models than Mr. Smith, and Mr. Fox’s students utilized the affordances of the virtual simulation more purposefully and in intended ways to think and reason about science content. Further, Mr. Fox’s students seemed to learn more during the unit about the affordances and constraints of models than Mr. Smith’s students, as indicated by better performance on an open-ended question about the affordances and constraints of models. This suggests that teachers’ mediation to help students understand the affordances and constraints of models may have been important to help students understand how to use the models as tools for scientific thinking. Given that the affordances of tools are not always obvious to the user (Norman, 1988) and that explicit mediation may be needed (Kozulin, 2003), differences in the teachers’ discussion about affordances may have differentially helped students see and understand how to take advantage of the intelligence built into the models. This offers a more nuanced understanding of prior findings that teachers’ discourse can play an important role in mediating students’ interactions with models to help them meaningfully use the affordances of models in science (Martin et al., 2017).

Overall, both teachers discussed science content and practices more than models, which is consistent with prior research (Gnesdilow et al., 2010) and the suggestion that teachers may not explicitly support students to understand the meaning and affordances embedded in tools (Kozulin, 2003). However, while the teachers’ proportion of explicit talk about the affordances and constraints of models was low in comparison to their talk about science content and practices, we found that even the small proportion of talk by Mr. Fox about the affordances and constraints of models may have had a big impact on his students’ understanding of and engagement with models. Thus, it may be that minimal, but purposeful and explicit, discussion about the affordances and constraints of models by a teacher can have important benefits for students to understand the complex process of scientific modeling. A productive avenue for future work could be to further investigate the utility of working with teachers on conducting such discussions with their students, or designing prompts about affordances and constrains of models within curricula and technologies such as the virtual simulation used in this study.

Another significant difference in the types of mediation that the teachers provided their students was that Mr. Smith engaged in more discourse focused on science content. It is not clear whether Mr. Smith’s discourse contributed to or was a result of his students’ use of the models. For instance, was it the case that Mr. Smith's students did not take advantage of the tools as well as Mr. Fox's students did, because Mr. Smith did not talk about the affordances and constraints enough? Or, was it the case that Mr. Smith spent most of his time discussing science content to redirect students’ attention from the virtual simulation to get them to think more deeply about the underlying science? Asking students questions related to science content may have been productive to engage students in scientific thinking in the moment while the teacher was present, but perhaps this type of mediation was not sufficient for supporting students to use the models to engage in scientific thinking when they were on their own. While Mr. Smith engaged in some deep conversations about science content with his students, our results suggest that this type of mediation may not be sufficient to help students understand and utilize models as scientific tools. As suggested by others, balancing support for science content with science practice, the practice of using models in this case, seems crucial for student learning (e.g., Lehrer & Schauble, 2006; Windschitl et al., 2008).

While Mr. Fox rarely discussed the nature of models overall, his talk about the affordances and constraints of the models seemed to help his students use them in ways that were intended. However, we do not know if this understanding would transfer to students’ use of other models or other contexts in the future. Since tool use is situated and context dependent (Kozulin, 2003; Pea, 1993), it will be important to further investigate what forms of teacher mediation are needed for students to be able to do this. For instance, we still do not know the extent to which explicit discourse about the general nature of models might further help students use a variety of scientific models effectively, since the teachers engaged in little of this type of mediation. Future research with more teachers in different contexts will be important in exploring these questions. Additionally, the fact the Mr. Fox’s class was a “challenge” class at his school with students capable of engaging in more open-ended activities may have contributed to his students needing less mediation than Mr. Smith’s students. Higher ability learners might be able to better take up mediation provided by a teacher or simply be able to better use the affordances of multiple tools. Future research could further investigate how teachers’ mediation might affect students of different abilities to use multiple models for scientific thinking.

Even though scientific models are pervasive in K-12 science classrooms, students often have difficulty using scientific models as tools for thinking and building knowledge without the help of a teacher (Schwarz et al., 2009). As multiple models are employed in future science classrooms, helping students to capitalize on the
unique affordances of virtual and physical models will be essential to support their learning. It is important that students understand how to use these types of models together as tools for scientific thinking and for teachers to know how to support this learning. It is well documented that students struggle to understand and utilize scientific models (Schwarz et al., 2009; Treagust et al., 2002) to learn. Yet, little research has investigated how students use both virtual and physical models in an integrated way to learn science (Olympiou & Zacharia, 2012), and even less research has investigated how teachers can mediate students' interactions with these models to support students' scientific modeling practices. Our paper begins to address this gap and highlights the important role of teachers’ discussions about the affordances and constraints of multiple models in helping students to understand and use models as tools for scientific thinking.

References

Acknowledgments
We thank the teachers and students who participated in this study. This research reported in this study was supported by a DRK-12 grant from the National Science Foundation to the last author.
Designing for and Analyzing Productive Uncertainty in Science Investigations

Eve Manz, Boston University, eimanz@bu.edu

Abstract: This paper explores methods to design and analyze learning environments in which uncertainty is incorporated into students’ activity to support the development of science practices and content understandings. I describe how a second grade landforms investigation was designed to incorporate productive uncertainty about the relationships between phenomenon, empirical investigation, evidence, and explanation. I then describe, apply, and reflect on an analytic method to trace uncertainty through (1) instantiation in the learning environment (2) recognition and introduction into conversation, and (3) conversations in which uncertainty supports participation in science practices.

A major question in current research in science education is how to engage young students in the epistemic practices of science in ways that are meaningful to them and powerful for the development of science understandings. Making progress on this challenge will involve (1) understanding what forms of practice are most appropriate targets for young students, (2) developing principles for designing learning environments that support meaningful engagement in practices, and (3) understanding the forms of support that allow teachers to orchestrate these learning environments.

Recently, scholars have stressed the importance of establishing a need for science practices in learning environments. Researchers have argued that unless students participate in science practices in the context of “figuring something out,” those practices are likely to have little scientific meaning (Berland and Hammer, 2012). They have stressed that learning environments should highlight disciplinary forms of uncertainty and allow students to grapple with them (Engle, 2011; Reiser, 2004). From this perspective, science practices are emergent from the learning environment, in the sense that students engage in and develop science practices for their own science activity (Cobb & Yackel, 2006). These approaches focus on the purposes of science practices, ask when those purposes are meaningful to young students, and use methods to analyze the local development of practices in classroom communities. This paper seeks to make a contribution to emergent approaches to science practices: first by exploring how forms of uncertainty that are under-represented in learning environments can be productively built into young students’ science investigations, and second, by developing a method for tracing the relationship between uncertainty and the emergence of science practices.

Incorporating uncertainty into elementary science investigations

The work reported here draws from socio-cultural and emergent approaches (Cobb & Yackel, 1996; Saxe, 2002) that treat practices as constituted and adapted in communities to solve shared problems. Scientific activity is driven by the need to manage uncertainty; uncertainty not only about how to explain the world, but how to represent the world in the form of an experiment, what to measure, and how to convince peers to see what the scientist wants them to see. These uncertainties manifest for scientists as decisions about what to do (e.g., what the best experimental design is), as push-back from empirical work (for example, in the form of surprising experimental results), and as critique from an audience (e.g., the disagreement of a peer about whether a measure is appropriate) (Knorr-Cetina, 1999; Pickering, 1995). For scientists, uncertainties constitute sites for argumentation, explanation, and the development of new understandings. From this point of view, students need to experience some of this uncertainty if they are to engaging meaningfully in scientific practices and develop an understanding of what those practices are for (Ford & Forman, 2006; Manz, 2015).

Productive uncertainty in classroom learning environments

When I refer to scientific uncertainty, I mean an aspect of scientists’ work that is non-obvious and contingent, which must be figured out by the scientist and negotiated in response to feedback from peers and the material world (Pickering, 1995; Rouse, 1999). I use the term productive uncertainty as a pedagogical construct to describe an approach where students might profitably engage with scientific uncertainty, in the sense that by grappling with some of the decisions scientists must make, students would make progress on scientific practices and consent understandings. Constructs that I draw on include research on productive disciplinary engagement that stresses the need to “problematicize” content (Engle, 2011; Reiser, 2004), the use of “cognitively demanding tasks” (Henningsen & Stein, 1997), and “productive failure” (Kapur, 2008). I draw from these literatures to identify three sets of principles for incorporating productive uncertainty in learning environments.
Understanding what to make uncertain
Students should engage in tasks designed to open up decisions or problems that are important from a disciplinary point of view, without so much ambiguity that the learner cannot engage with key uncertainties.

Helping students experience, and make public, moments of uncertainty
Students must recognize and engage with disciplinary uncertainties: there must be a moment where students feel puzzled, see a conflict, or disagree with a member of the classroom community. These moments establish a need to engage with disciplinary practices and content.

Helping students resolve or learn from uncertainty
There needs to be support to resolve or learn from the uncertain experience, in the form of tools that scaffold students’ engagement and teachers guiding discussions where strategies and solutions are compared, and, when necessary, introducing and unpacking next steps or canonical procedures and ideas.

Empirical uncertainty as an under-leveraged resource in elementary classrooms
Drawing from the Science Studies Literature (e.g., Gooding, 1990; Pickering, 1995), I conceptualize important forms of uncertainty as emerging for scientists as they manage the transitions between the complex and material phenomena they seek to understand, the empirical investigations that they use to “get a grip on” the world, the observations and evidence that the investigations generate, and the explanations and explanatory models of phenomena that are the targets of scientists’ work (Manz, 2015). The default assumption, represented in many curricula and research-based interventions, is that the easiest entrée to scientific practice is to ask students to engage in highly simplified investigations so that they can learn to support claims with evidence, then later problematize what counts as evidence. The result is that forms of uncertainty that are central to scientific work are often obliterated or made invisible in elementary science investigations. For example, elementary students rarely are asked to consider how to design an experiment to represent a phenomenon, determine what to count as evidence, or consider misfits between the experiment and phenomenon as they apply their results to understand the world. Several studies have demonstrated that young students can recognize these forms of uncertainty and productively grapple with them, but we know little about how to systematically design, and support teachers to orchestrate, environments that use empirical uncertainty to support scientific practices and understandings.

Central conjecture
Drawing from the three principles for productive uncertainty described above, I conjecture that forms of empirical uncertainty that are typically removed from elementary students’ experience with experiments can be productive for students’ development of scientific practices and content understandings. However, I also conjecture that for uncertainty to be productive, it must be strategically designed into the learning environment and carefully managed by teachers.

In particular, the conjecture that drives this study is that engaging students in a rich phenomenon and choice in how to investigate it can generate variability in students’ methods, claims, and evidence; this variability can in turn be recognized and made public by students as they present and compare findings, in turn eliciting sense-making discussions in which teachers support students to engage in science practices and consider important content understandings. In previous work (Manz, 2015), I have explored this conjecture in one instantiation of students using a plant growth experiment to understand plant needs and growth patterns in a wild backyard area. In this study, I sought to apply this conjecture to working with a small group of teachers, in order to better understand the design elements in the learning environment and teacher discussion moves needed to support the implementation of productive uncertainty in empirical investigations. Below, I describe the context in which I enacted the study and the analytic methods developed to further test and refine the conjecture.

Study Context
This work was conducted in a suburban district in the Northeastern United States that had recently adopted new materials to better align to the state’s new standards, which are modeled on the Next Generation Science Standards (Achieve, 2013). The district elementary science leader approached the researchers to help her adapt the curriculum materials to support deeper sense-making opportunities for young students. She recruited five second-grade teachers to work with the research team on adapting one of their investigations. The teachers varied in their years of teaching experience from 2 to more than 15 years of experience, in their science content understandings, and in their comfort teaching science, as is typical of elementary teachers in the United States.
The research team first supported the teachers to conduct an investigation that involved scientific uncertainty (in the sense that teachers had to make decisions about how to represent the phenomenon, what to use as evidence, and how to interpret their findings) and debrief how the need to make decisions and the different decisions made supported sense-making, explanation, and argumentation. We then worked with teachers to adapt an investigation from the science kit used in their district to incorporate scientific uncertainty as an opportunity for student sense-making. The focal investigation comprised two lessons from a landforms unit. In the investigation, students sprayed soil, sand, and gravel with water and blew on the materials through a straw to understand how wind and water can shape land. The kit did not provide direction for teachers to support students to think about how to represent wind and water, how to design an informative comparison, what data to collect, or how their findings allowed them to understand how wind and water might shape land.

We worked with teachers over five meetings to create a shared set of lesson plans that spanned approximately six 45 minute lessons. In the re-designed lessons, students examined photographs and discussed how wind and water might shape land by moving earth materials; examined earth materials and made predictions; designed investigations in groups using a straw and spray bottle to test their ideas; developed and supported claims in small groups; presented and critiqued claims and evidence; and discussed the phenomenon again based on their investigations. Table 1 shows how the sets of design principles for incorporating productive uncertainty in learning environments were instantiated in this design.

Table 1: Productive uncertainty design principles as instantiated in the landforms investigation

<table>
<thead>
<tr>
<th>Design principle</th>
<th>As instantiated in this design</th>
</tr>
</thead>
</table>
| What to make uncertain                                | - Focus on allowing students to experience uncertainty about how to represent a phenomenon in an experiment, how to construct evidence, how to interpret evidence, and how to use their investigation to explain the phenomenon.  
  - Bound uncertainty so that students make a limited number of choices (e.g., give students materials to use but allow them to decide how to use them)  
  - Students design experiments in small groups                                                                                                                  |
| Helping students experience and make public moments of uncertainty | - Allow student groups to come to different conclusions about whether and how wind and water move the earth materials  
  - Student presentations and student questions after experiments are completed  
  - Examining original and new phenomena after experiment completed: share examples of movement (e.g., boulder in a field) that cannot be explained by investigation findings.                                                                                       |
| Helping students resolve and learn from uncertainty   | - Teacher facilitation of discussion of differences in findings  
  - Beginning and ending set of lessons with the same phenomenon                                                                                                    |

Methods
A team of two researchers collected data over the course of the investigations, from the initial introduction of the phenomenon to the discussion of the phenomenon after students’ experiments were concluded. Four teachers consented to videotaping. In these classrooms, one video camera was used to film whole group discussion and student groups during small group work. All classroom artifacts and student work were collected. In the fifth classroom, researchers attended all lessons, took close field notes, and photographed artifacts.

We next developed a method to test and refine the conjecture, which we represented as:

Rich phenomenon + Choice in how to represent and investigate it → Variability in students’ initial ideas, methods, claims, and evidence → Recognition and discussion of variability → Development of practice and conceptual understandings.

This paper focuses on our initial development and application of methods to analyze the instruction in two of the classrooms. We chose to begin with these two classrooms because we had complete records for them and because the teachers were the most likely to allow students to make decisions and then to discuss those decisions. Therefore, these data sets provided the richest opportunity for developing and refining analytic methods. Because of the short time frame of the investigation, the data was used to support inferences about
opportunities to participate in practices and consider content, rather than to detail development. Data analysis occurred in four phases.

**Phase 1: Activity map**
In the first phase, we used field notes and classroom artifacts to create an activity map for each classroom that summarized the major shifts in activity and the video, artifacts, and student work associated with each activity. For example, activities for both classrooms during the first two days were discussing pictures of the phenomenon, examining earth materials and making predictions about whether wind and water can move them, whole group predictions, introducing the investigation, and small groups planning investigations. We organized the activities into three major stages: (1) Examining the phenomenon and making predictions, (2) Planning, conducting, and making claims in small groups, and (3) Sharing findings and discussing implications.

**Phase 2: Cataloguing and summarizing activity**
In this phase of analysis, we moved systematically and sequentially through all video and classroom artifacts for each classroom investigation. We developed focal categories for each stage of the investigation, independently coded and refined categories based on a sample of the data, then exhaustively sampled and coded the data for all episodes or artifacts related to the focal categories.

**Phenomenon and predictions: How did students make sense of the phenomenon?**
For this stage, our focus was describing the aspects of the situation that students thought were relevant as they examined the phenomenon and the earth materials. These included: the size of the particles, the strength of the wind or water, explanations for how and why materials move (e.g., water makes soil heavier and harder to move, water loosens dirt and then allows rocks to move). In addition, the predictions, and variability in predictions, made by the class were summarized (e.g. all students agreed that wind could move sand, students disagreed about water moving sand because some thought that sand might clump and not move).

**Conducting investigations: What forms of uncertainty were realized in the learning environment?**
For this stage, we focused on what forms of scientific uncertainty were realized in the learning environment. Following the study conjecture, we defined focal forms of uncertainty as decisions that students made about how to represent a phenomenon in an experiment, how to construct evidence, how to interpret evidence, and how to use investigation to explain phenomenon. As indicators of students making decisions, we documented teacher and student questions, variability in how groups made decisions, and evidence of disagreement and discussion during small group planning. For example, in each classroom, we documented the following forms of uncertainty: the angle and distance that the straw was held in relation to the materials, what to use as evidence of movement, and what claim to make.

**Discussion: What forms of uncertainty were brought up and how were they addressed?**
In this investigation stage, students and teachers presented and discussed findings, agreed on a conclusion about the investigation, and discussed the implications for understanding phenomena in the “real-world.” We documented how the forms of uncertainty in Stage 2 were introduced by teachers and students in these discussions and analyzed when and how the introduction of the uncertainty supported moments where students engaged in science practices and explored new or deeper science content.

**Phase 3: Tracing uncertainty through each investigation**
In this third phase of analysis, we developed a summary table for each classroom in which we followed each form of uncertainty that students and teachers grappled with through the course of the investigation, with an eye toward understanding: (1) how the uncertainty was realized in the learning environment, (2) how it was introduced into whole group conversation, and (3) whether and how it supported opportunities for whole-group sense-making. Table 2 shows one row from this table for one teacher, following decisions generated around the strength of the spray of the water bottle (as operationalized in distance the water bottle was held from the materials, nozzle setting, and amount of water sprayed). Each sub-row refers to one instance of strength of spray that was brought up in whole group conversation.
Table 2: Example row for forms of uncertainty summary table: strength and amount of water in Ms. A’s class

<table>
<thead>
<tr>
<th>Form of uncertainty</th>
<th>Evidence that we see the uncertainty realized</th>
<th>How it was brought up in whole class discussion</th>
<th>What happened</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength of water sprayed by the spray bottle (How to operationalize the phenomenon in an informative comparison)</td>
<td>Different groups choose different numbers of squirts in their investigations. There is talk about strength of spray in groups (though nozzle setting is set in Ms. A’s class). Students seem to be using different distances in video of groups performing investigations.</td>
<td>170526_V3 [00:07:53.27] During a group’s presentation, student asks if the presenters planned the number of sprays they used. 170526_V3 [00:09:17.27] Researcher points out that class disagrees about claims. 170526_V3 [00:12:12.23] When students are asked to generate a whole-class experiment, a student points out amount and strength of sprays as something to control. 170526_V4 [00:02:13.08] Researcher asks why it matters that sand and soil get the same number of sprays.</td>
<td>When group answers five, Tr. directs students to identify that this controlled the amount of water across conditions. Student notes that they might have different ideas because some groups sprayed harder. Tr asks why that’s important and evaluates the idea as important to experimental design. Student answers with control of variables response – experiment doesn’t work the same if you change the sprays, obviously the sand would move farther. Tr. briefly asks other students if they agree. No student response invited. A student says that in her experiment, they did rocks last and maybe the bottle didn’t spray as much.</td>
</tr>
</tbody>
</table>

Phase 4: Developing themes
In this phase, the two researchers separately reviewed the summary forms for each of the two teachers and developed memos to summarize the forms of uncertainty that emerged in each classroom, the ways that these were introduced into whole group conversation, and the kinds of conversations that emerged. We were particularly interested in understanding whether students or teachers were bringing up uncertainty, how teachers responded to different forms of uncertainty, which forms of uncertainty appeared to be productive for science practices and content understandings, and challenges that teachers and students faced. We then worked together to compare, discuss, and refine themes.

Findings
Table 3 shows the forms of uncertainty that were realized in the learning environment, and the ways that these were introduced into classroom conversation. I first describe these columns more fully. I then share findings about how the introduction of these forms of uncertainty supported conversations that allowed students to engage in science practices and explore or deepen content understandings, and document challenges that teachers and students faced. We then worked together to compare, discuss, and refine themes.

We documented similar investigation choices in each classroom; however, the choices that students grappled with differed somewhat across the two classrooms. These differences appeared to be related to teacher choices about how to introduce and guide the investigation. In each classroom, students made different choices in how they positioned the straw and water bottle, how they arranged the earth materials (e.g., leaving them in the petri dish, placing them in a mound out of the dish), and what they used as evidence of movement (e.g., a qualitative comparison of distance, measuring distance, separation in the petri dish, materials floating).
However, we also saw differences in the choices that were realized in students’ investigations across classrooms. For example, in both classrooms, our videotape records provided evidence that at least one group considered using the water bottle to make a pool of water. However, in Ms. B’s classroom this strategy was not taken up; in fact, a classroom assistant working with the group redirected them to use the bottle to spray. In Ms. A’s class, a researcher intervened to make space for and amplify a student’s proposal to use the spray bottle to fill the petri dish, then put the materials into the dish and see how they moved when the water was moved. In contrast, in Ms. B’s class, methods for marking and measuring distance were highlighted and encouraged, and a greater variability in measurement strategies was generated.

Table 3: Forms of uncertainty realized in the learning environment and introduced into whole class discussion

<table>
<thead>
<tr>
<th>Uncertainty</th>
<th>Class</th>
<th>Ways introduced into whole class discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>How wind and water shape land</td>
<td>Ms. A, Ms. B</td>
<td>Students make different predictions about earth materials and draw on different experiences; teachers highlight disagreement.</td>
</tr>
<tr>
<td>What to use to represent wind &amp; water</td>
<td>Ms. A, Ms. B</td>
<td>Teacher asks students to consider, then introduces and provides rationale for decision to use spray bottle and straw.</td>
</tr>
<tr>
<td>How to use straw: strength</td>
<td>Ms. A, Ms. B</td>
<td>Teacher highlights as something to consider, students bring up spontaneously when asked how they are planning as something to consider and control, students bring up as something to question each other about, students bring up as accounting for different results, student brings up as something that the experiment doesn’t represent well about the real world.</td>
</tr>
<tr>
<td>How to use straw: angle/direction</td>
<td>Ms. A, Ms. B</td>
<td>Teachers highlight as something to consider.</td>
</tr>
<tr>
<td>How to use spray bottle: form of water (rain vs. pool)</td>
<td>Ms. A, not fully in Ms. B.</td>
<td>Presenting students use different methods, students question each other about method, teacher introduces alternative (pool) as a method for whole class, students disagree about how alternative can generate evidence.</td>
</tr>
<tr>
<td>How to use spray bottle: strength</td>
<td>Ms. A, Ms. B</td>
<td>Teacher points out as something to consider, students question each other, students generate as explanation of different results, students bring up as something to consider when planning, teacher draws attention to differences in joint experiment.</td>
</tr>
<tr>
<td>How to use spray bottle: angle/direction</td>
<td>Ms. A, Ms. B</td>
<td>Students ask questions, students bring up to explain different results, student disagreement about angle to use in joint test.</td>
</tr>
<tr>
<td>Earth materials in vs. out of dish</td>
<td>Ms. A, Ms. B</td>
<td>Teacher brings up as something to consider and ask about; Students bring up as not representing world.</td>
</tr>
<tr>
<td>How materials are arranged</td>
<td>Ms. A, Ms. B</td>
<td>Students bring up as something to ask each other about; students use to explain different results (clumping materials, big vs. small rocks).</td>
</tr>
<tr>
<td>Order wind and water are tested</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Form of evidence generated for wind</td>
<td>Ms. A, Ms. B</td>
<td>Students introduce different evidence into conversation, students ask each other about evidence, students ask to define what counts as evidence (e.g., what counts as moving), teacher introduces and ratifies particular forms of evidence (measuring distance), students disagree about evidence in joint investigation.</td>
</tr>
<tr>
<td>Claims about wind</td>
<td>Ms. A, Ms. B</td>
<td>Students introduce different claims into whole group discussion, students point out disagreement in claims and generate explanations for disagreement.</td>
</tr>
<tr>
<td>Form of evidence generated for water</td>
<td>Ms. A, Ms. B</td>
<td>Teachers ask about evidence in recaps, teachers ask student presenters about evidence, students introduce different forms of evidence into conversation, students ask each other about evidence and clarify evidence, students spontaneously compare evidence.</td>
</tr>
<tr>
<td>Claims about water</td>
<td>Ms. A, Ms. B</td>
<td>Students introduce different claims into whole group discussion, students note that their claims were different from other groups and seek to explain.</td>
</tr>
</tbody>
</table>
In the two classrooms, investigation choices were introduced in whole group conversation in ways that positioned students as making and justifying decisions; that is, as grappling with scientific uncertainty. First, teachers highlighted differences in what students were doing, probed student methods, and in one classroom asked students to make and justify decisions to develop a joint experiment. Students were positioned in these conversations as navigating uncertainty about the angle and distance of the spray bottle and straw, how they arranged materials, what they paid attention to in order to determine whether the earth materials moved, and what claims to make. In addition, these forms of uncertainty were introduced into conversation by students when asked to talk about investigation decisions, suggesting that students were taking them up as legitimate decisions that required sense-making. Importantly, we also noted that students brought up these decisions to account for different results, to question their classmates about, or to disagree with methods proposed by others. For example, students in Ms. A’s class quickly suggested that they likely used the straw differently after they participated in a gallery walk that made visible that they had come to different conclusions. In addition, when groups presented their claims and evidence, students asked each other about the angle and strength of blowing and spraying, clarified how far different materials had moved, and (most rarely) explicitly disagreed that a method could be used to support a claim, as when a student disagreed that moving the materials in the petri dish showed that they moved with water, because the tester was moving the dish, and therefore moving the materials.

Analysis of the discussions demonstrated several teacher responses to the introduction of uncertainty. One response when teachers heard variability in decisions that might affect results was to ask questions to probe what students had done and make differences visible to other students. Another was to highlight an idea that a student brought up (for example, asking to compare the nozzle setting used by another group) as important, perhaps noting it on the board as something to consider or ask others about, then moving on. A third response was to open up a question or disagreement to talk by multiple students. A fourth was to effectively dismiss or argue against the concern: this was rarer, but informative, as it was most likely to occur in one classroom and at times where the teacher was focusing on bringing the class to a joint conclusion, for example about how to conduct their joint investigation or to develop a conclusion that wind and water can move rocks, sand, and soil.

Finally, we closely analyzed conversations to understand what opportunities they provided for students to participate in science practices and develop content understandings. Analysis suggested that when teachers made space for student questions and for multiple turns of student talk, students had opportunities to engage in practices related to investigations, in that they made decisions about and brought up the importance of controlling variables, of considering the level of variables used in their experiments (e.g. the strength of the force of the wind or water), and of considering how an investigation did and did not represent the phenomenon. They engaged in practices of argumentation they considered why their claims were different and probed each other’s evidence. In addition, the investigation provided opportunities for students to engage in explanation as they explored the reasons for their results, tried to understand differences in results (for example, there was a lengthy conversation in Ms. B’s room about why a downward spray might exert force that moved rocks more than sand), and explored why the investigation could not fully help them explain phenomena of wind and water changing land, often proposing new mechanisms for movement or making new connections to phenomena.

We also noticed several challenges across both classrooms. First, during presentations, teachers often interjected so often that students were actually able to ask each other relatively few questions. Second, we noted that in discussions teachers only very rarely helped students relate the forms of uncertainty that they were recognizing to the goal of trying to agree on a shared conclusion about the experiment or understand the phenomenon. That is, they were more likely to name something to consider (“We have to control that,” “you’re thinking about the real-world”) than to ask “Why would that matter?” or “Could that explain why…” Finally, we noted that students’ consideration of ways that the experiment might not represent the phenomenon of wind and water shaping land appeared both to be particularly rich for the development of explanation and content understandings, in that students were beginning to make visible the mechanisms by which wind and water shape land, as well as particularly difficult for teachers to take up and guide, especially in the cases that these considerations were at odds with either controlling variables or using evidence to support a shared conclusion.

Discussion
The conjecture that drove this work was that incorporating forms of empirical uncertainty that are typically left out of elementary students’ empirical investigations can generate variability in ideas and methods that, when recognized and introduced into classroom conversation, can support productive episodes of sense-making in which students engage in scientific practices and make progress on content understandings. In this paper, I described and applied an analytic method for cataloguing and following uncertainty to explore this conjecture.

The analytic method is one that our research team is finding productive. In particular, we are finding it useful to separately consider and relate the three related aspects of (1) uncertainty as realized in the learning process, (2) the way uncertainty is discussed and negotiated in classroom conversation, and (3) the way uncertainty is made visible to and considered by students as part of their engagement in scientific practices.
environment through variability in student choices and thinking, (2) uncertainty as recognized by students and introduced into conversation as a decision that requires sense-making and that might account for differences in results, and (3) uncertainty as grappled with in conversation, potentially supporting science practices and sense-making about content. Here, engaging in these three forms of analysis allowed us to document rich variability in student decisions, show that students could recognize the importance of investigation decisions and initiate conversation about them, and explore opportunities and challenges in the ensuing conversations. I expect that all three forms of analysis will be essential to generating learning environments where uncertainty is productive.

We plan to more closely analyze the investigation as taught in all five classrooms to understand which forms of uncertainty appeared most productive and which appeared at odds with each other (e.g. a focus on controlling variables and representing the phenomenon of wind and water shaping land) to inform the next iteration of designing the investigation. We will then refine what is made uncertain (productive uncertainty design principle 1 above) to focus attention on those decisions that appeared to most productively support sense-making and to help teachers attend to and draw out those uncertainties as a resource for science practices and understandings. I suspect that such a step; that is, generating more opportunities for uncertainty in the first instantiation in a design in order to understand which offer the greatest potential for student learning, is a fruitful move to make in designing for science practices and understandings.

In addition, the similarities in the decisions that students grappled with suggest that certain forms of uncertainty might be predictably generated and recognized by students in this investigation; this finding will be further tested with the remaining three teachers, and if stable, could support the development of an investigation that is legitimately uncertain for students but somewhat predictable for teachers. If this is the case, we could better draw from practices used in mathematically cognitively demanding tasks, where teachers anticipate, select, and juxtapose different solution strategies (Stein, Engle, Smith, and Hughes, 2008) to support productive discussion. Differences across instantiations (e.g., that in one classroom, where the teacher used gallery walks and highlighted disagreement, student were more likely to bring up variability in methods to account for findings) can support the refinement of design principles for productive aspects of the learning environment and teacher moves that elicit and support uncertainty. Similarly, the consistent challenges that we noted, e.g., teachers not discussing the implications of surfaced forms of uncertainty, can be used to refine conjectures and methods of working with teachers, for example, by developing teacher moves. This work can support the important goal of moving past cookbook elementary school investigations while developing the supports that teachers and students need to manage this more complex instantiation of scientific work.

References
Public Peer Review Motivates Higher Quality Feedback

Xu Wang, Carnegie Mellon University, xuwang@andrew.cmu.edu
Yali Chen, Cengage, Carnegie Mellon University, yali.chen@cengage.com
Amanda Godley, University of Pittsburgh, agodley@pitt.edu
Carolyn Rosé, Carnegie Mellon University, cprose@cs.cmu.edu

Abstract: The role of feedback in learning has been well researched, but in practice high quality feedback may be scarce, for example when the source of feedback is from peer learners. Nevertheless, peer feedback may be the main source of formative feedback available in some settings, such as in Massive Open Online Courses (MOOCs). A key part of the problem may be that students do not have sufficient incentive to offer their best feedback in settings where supervision is minimal. In this paper, we investigate whether students provide feedback of higher quality when it is done in a public setting rather than in a private setting. We report on an experimental study with 65 participants randomly assigned to a public feedback and a private feedback condition. We report the effect of the manipulation in terms of the quality of feedback offered as measured by a validated coding scheme, the subjective rating of the feedback, the effect on propensity to revise and success at increasing the quality of the writing. Limitations of the study and implications for practice are discussed.

Introduction

Much research in the Learning Sciences has investigated the properties of effective feedback and the way it plays into the learning process. For example, Graham et al. (2015) found that feedback significantly enhances students’ performance on writing; The Knowledge-Learning-Instruction framework (Koedinger et al., 2012) considered timely feedback to be a critical instructional activity across disciplines, from English language learning to science and math; The ICAP framework (Chi & Wylie, 2014) classified peer feedback as an interactive activity, which is a category of activity designated as particularly conducive to learning. Despite the importance of feedback for learning, lack of available high-quality feedback is sometimes the norm, for example in at-scale online learning environments such as Massive Open Online Courses (MOOCs) (Kulkarni et al., 2015; Hicks et al., 2016; Joyner et al., 2016). Research related to the improvement of peer review processes has not always leveraged best practices in feedback research from the Learning Sciences or measured success in terms of qualities of feedback known to be important for learning. On the one hand, questions have been raised regarding a lack of student motivation in writing feedback, resulting in feedback of inconsistent quality (Suen, 2014). However, it is an open question of whether increasing motivation alone would increase the quality of feedback provided by students when feedback is evaluated using best practice rubrics grounded in the Learning Sciences. In this paper, we address the practical question of how to increase availability of high quality feedback by testing the effect of a public peer review paradigm in at-scale learning environment. We measure the impact of the manipulation on the quality of feedback offered, the subjective rating of the feedback, the effect on propensity to revise and success at increasing the quality of the writing. In this way, we take a first step towards bridging work in peer feedback from practice-oriented communities with basic research on feedback and writing from the Learning Sciences. The long-term goal is to leverage students in online learning settings as a resource for one another. Massive online learning settings pose challenges, but with proper scaffolding the scale may eventually serve as an advantage rather than an impediment.

Peer review has become increasingly popular in MOOCs since instructors are not able to grade and provide feedback for the large number of assignments turned in by students. However, in a near anonymous online learning environment, there are few consequences for high or low-quality feedback and the social distance makes it less likely that students are willing to invest substantial effort in offering feedback. Similarly, prior work has found that when forms of discussion identified as valuable for learning in offline settings occur in MOOCs, they are associated with learning as expected, but rarely occur without support (Wang et al., 2016). On the other hand, Wen et al. (2018) found that feedback exchanges in a public discussion forum prior to separation into collaborative groups is more valuable as preparation for collaboration than feedback delivered privately within the collaborative groups once they are formed. This result in particular suggests that in a public forum, either students are motivated to give better feedback or that students benefit from exposure to a wider variety of work and feedback in a more public context. Tinapple et al. (2013) also points to the importance of accountability when providing peer feedback, which may also be a key factor in a public environment.
In traditional peer review scenarios, which are typically more private, students’ exposure to classmates’ work and feedback is limited to those from the few classmates they have been matched with. Similar to Wen et al. (2018), we explore a new peer review paradigm, where students post their work to a public discussion forum, and then write feedback to a small number of other students in the forum, thus getting full exposure to all assignments and reviews in the class. We investigate whether the increased accountability of the public environment results in higher quality feedback being exchanged when scaffolding is introduced to focus the effort on forms of feedback known to be beneficial. Prior work deconstructing types of feedback (such as summary, praise, problem, solution, etc.) has found specifically that summarizing the peer’s ideas, identifying problems and offering solutions in feedback on writing is associated with better revision, and thus considered to be preferable characteristics of feedback (Nelson & Schunn, 2009). In our study, we specifically point students to these aspects of feedback in our instruction and investigate whether students in a public setting are more conscientious of the instruction provided, and go on to provide more high-quality feedback. This study contributes new knowledge about which aspects of feedback (namely, summary, praise, problem and solution) can be manipulated through raising the level of perceived supervision as a result of situating the feedback activity in an apparently public setting. The study also reveals the extent to which raising the prevalence of the preferable aspects of feedback (such as solution) could contribute to revision of writing in a near-anonymous public review environment. The results challenge us to probe deeper in order to identify strategies that will achieve substantial positive impact in practice.

Theoretical foundations and hypotheses

An increasing number of interventions are being designed and developed for peer grading in MOOCs (Kulkarni et al., 2015; Hicks et al., 2016; Joyner et al., 2016), ranging from novel grade aggregation techniques (Sajjadi et al., 2016), to enabling students to rate their graders in return (Staubitz et al., 2016). Prior work on peer assessment in MOOCs mostly focused on getting an accurate grade to students. The value for learning and improvement of performance as a result of feedback has been far less of a focus. Some work on infrastructure for supporting the feedback process in online settings provides a practical foundation for our work. For example, PeerStudio (Kulkarni et al., 2015) enables students to provide feedback to their peers in MOOCs. A class survey shows that students found free-form comments to be more useful than the grade, and students liked reading each other’s work the most, more than getting feedback and revising their work. It was found in Joyner et al. (2016) that meta-reviewers wrote higher quality feedback based on the first round of reviews. In Tinapple et al. (2013), some comments from students showed how reviewing their peers’ work increased accountability and a sense of community. For example, “You take the work into account much more when you are aware that your peers will be reviewing it” and “This class had a sense of community which I think many college courses lack. When someone feels more comfortable in the class, they will do better in that class.” Staubitz et al. (2016) found that students had increased motivation to provide high quality reviews when they had the opportunity to rate their reviewers.

Such prior work has shown the potential for positive impact when giving students access to more reviews. Our work points to the accountability issue where students may feel more obliged to provide high quality feedback when their feedback is disclosed to the class instead of provided to individuals privately. On the other side, prior work has also shown that a lack of sense of community could give rise to feelings of disconnection and affect students’ persistence (Kerka, 1996), while emphasizing community reduces drop out (Tinto, 1993). Exchanging feedback in a discussion forum may provide this needed sense of community.

There has been a line of research in the Learning Sciences that studies what makes feedback effective. For example, an authoritative reference is a coding manual for feedback quality developed in widely acknowledged work of Nelson & Schunn (2009). This coding manual was informed by a survey of prior work on what kinds of feedback led to better performance outcomes. For example, Ferris (1997) found that feedback that included summaries promoted more substantive responses to feedback. The same work also reported that specific comments were more helpful than general comments. Bitchener et al. (2005) found that feedback that contained solutions was more helpful for adults’ writing performance. We use the framework contributed by this body of research as a basis for both scaffolding and assessment of feedback in our study.

Based on this review of prior work, we propose the following two hypotheses. (1) We hypothesize that feedback generated in a public environment that includes scaffolding for offering effective feedback will be of higher quality than feedback generated in an otherwise equivalent private environment. In addition to the effect on feedback quality, we will also examine whether feedback received in the public environment actually helps students improve their writing. (2) We hypothesize that students who receive feedback in a public environment will have a higher revision rate and show a higher quality of revision than students who receive feedback in a private environment.
Method

We tested our hypotheses in a high internal validity setting as preparation for later research in high external validity large-scale online learning settings. We adopted a between-subjects design with random assignment in which we manipulate whether students write peer review in a public discussion forum or privately on a webpage. The study was run as a lab study in an online crowdsourcing environment, as we describe below. In both conditions, participants complete a 4-step task in which they write individual assignments and provide feedback to one other student’s assignment. In order to measure the quality of student feedback, we adapted the coding manual used in Nelson & Schunn (2009) and Patchan et al. (2016), and manually coded the feedback generated in the experiment into 9 constructs. We also developed a reliable coding manual to operationalize the quality of student assignments as control variables in our subsequent analyses. We use the same coding manual to measure the quality of student revisions. In the following subsections, we will provide a detailed description of the task, the coding manual for feedback, and the coding manual for assignment and revision quality.

Task description

All participants in the study participated in a 4-step task framed as a unit in an environmental sciences course, which took approximately an hour. In step 1 (~10 mins), students were asked to read learning materials on 4 types of energy. In step 2 (~20 mins), students were asked to write an energy proposal for a city. In step 3 (~15 mins), students were asked to provide a review to one other student’s proposal based on the scoring rubrics provided to them. In step 4 (~10 mins), students received feedback on their proposal, rated the feedback on how helpful they perceived it to be on a scale of 1-5, and then were offered the option of revising their own proposal. The experimental manipulation took place in step 3. In the private condition, students wrote their review on a webpage, whereas in the public condition, students were directed to a discussion forum and wrote the review using a similar interface and the same instructions but housed within the forum that was open to all participants. In both conditions, we introduced a feedback prompt. The feedback prompt works as a rubric that asks students to evaluate the quality of assignments from three perspectives, namely, 1) Clarity of arguments, 2) Supporting evidence, and 3) Adequacy of arguments as shown in Figure 1. Each piece of feedback thus contains three Comment Segment corresponding to the rubric. In the feedback prompt, we also provide guidance to students on how to write high-quality feedback, by specifically pointing to the preferable constructs of feedback as studied in prior work, such as “please summarize”, “be specific”, “give potential fixes”, etc.

Feedback quality coding

Our unit of analysis for feedback coding is Comment Segment. We adapted the first version of our coding manual from Nelson & Schunn (2009). In addition to the existing feedback constructs, we added Localized Praise. Though prior work has shown that Praise in feedback is at best ineffective and often diminishes future performance (Kluger & DeNisi, 1996), we might expect localized praise to be more helpful in preserving confidence than general praise. We also introduced a distinction in specificity for Problem and Solution, informed by Patchan et al. (2016), to distinguish feedback that focused on literal language usage (at a writing specificity level) and feedback that focused on logic or arguments (at an idea specificity level). Two researchers iterated on the definitions in the coding manual until sufficient inter-rater reliability was achieved over unseen data. In the final iteration, inter-rater agreement was evaluated for each category on 27 Comment Segment with final Kappa displayed in Figure 2. Once agreement was established, one of the researchers coded the whole set of feedback, blind to condition. A brief version of our coding manual is shown in Figure 2. The categories in the coding manual are not mutually exclusive; a comment could fall under multiple categories.

Subjective rating of feedback

Students are asked to rate how helpful the feedback comment is on a 1-5 scale, with 5 being “Very helpful” and 1 being “Very unhelpful”.

---

*Hi, welcome! Please provide feedback to your peer's assignment using the following rubric. Keep the two general guidelines in mind. Provide specific comments: point to exact places that were problematic; give examples, etc.*

1. Clarity of arguments: Please check whether the writing flows smoothly so you can follow the main argument. In your comments, please first summarize what you perceived as the main points being made so that the writer can see whether the readers can follow the paper's arguments. Then make specific comments about what problems you had in understanding the arguments and following the flow across arguments.

2. Supporting evidence from the readings: Please check whether the author's arguments are supported by evidence. If you found points made without support, describe which ones they were. If the support provided doesn't make logical sense, explain what that is. If some obvious counterargument was not considered, explain what that counterargument is. Then give potential fixes to these problems if you can think of any.

3. Adequacy of arguments: Please check whether enough perspectives or tradeoffs are considered. If you can think of any other factors that could play a role in a city's energy running, it is always a great idea to include different perspectives and provide extra information.

Figure 1: Feedback Prompt
We also coded the number of arguments, unsupported arguments and tradeoff / comparisons in the revision.

Participant recruitment

Assignments from the same three aspects. For clarity of arguments, we drew from the writing rubric from Purdue University (2017) and assigned scores of 1-4 to represent the clarity level (1: Beginning, 2: Developing, 3: Proficient, 4: Mastery). To evaluate the quality of supporting evidence, we counted the number of factors (knowledge points) students mentioned in their proposal, from a list of factors including budget, waste disposal, carbon emission, tax credit, wildlife protection, water quality, etc. If the author wrote arguments that were contrary to the information provided in the learning materials, e.g., “Coal energy is environmentally friendly,” we counted that as an argument without supporting evidence. Using the above standards, we operationalized the extent of supporting evidence in student writing in two categories, number of Arguments mentioned, and number of Unsupported Arguments. To evaluate the adequacy of arguments, we counted the number of tradeoffs or comparisons that the author has mentioned in the writing.

Using this coding approach, we formalized the evaluation of each assignment into 4 numeric variables, namely: Receiver Clarity: a score in the range 1-4; Receiver Number of Arguments: a score in the range 1-11 in our data; Receiver Number of Unsupported Arguments: a score in the range 0-3 in our data; and Receiver Number of Comparison/Tradeoffs: a score in the range 0-5 in our data.

After students had the opportunity to make revisions as part of the task, we computed the difference in their text between their original proposal and the revised proposal. We then used the same coding manual to code the quality of revisions. If the revision addressed a clarity issue, we coded revision clarity as 1, otherwise 0. We also coded the number of arguments, unsupported arguments and tradeoff / comparisons in the revision.

Participant recruitment

We ran the study as a lab study online through Amazon Mechanical Turk crowdsourcing platform from October to November in 2017. We ran the experiment in batches, with each batch associated with one or the other condition of the peer review environment. Therefore, each batch was assigned either to the public peer review condition or the private peer review condition. We restricted the participants to have above 98% acceptance rate in their work history and be located in the US. We did not collect information about participants’ demographics and education history, since their prior knowledge on the task would be accounted for by their assignment quality. The on-task time was ~55 minutes for each participant, with a possible 15-minute wait period. We compensated each participant $6 for their participation. We will further discuss the generalizability from crowdsourcing platforms to real MOOCs in the discussion section. For the public condition, we pre-populated the discussion forum with some existing student assignments. There were 3-7 people in each batch. They were
matched to provide feedback to each other’s assignments. In the end, we recruited 65 participants, with 32 in the private condition, and 33 in the public condition. We only included participants who completed the entire task. Though participants were told to write review to only one student, 3 students in the public condition wrote one more review voluntarily. This resulted in a total number of 68 pieces of feedback, with 204 Comment Segments.

Results

Hypothesis 1A: Investigating differences in feedback quality

In order to test our first hypothesis, we operationalized feedback quality in two ways, expert coding and subjective rating as introduced in the methods section. In this section, we present results regarding the effect of our intervention on the quality of feedback indicated by expert coding. We built regression models to compare feedback quality between the public and private conditions. Since we operationalized feedback quality into 9 constructs, we built 9 models to compare the quality difference for each feedback construct respectively.

For each feedback construct, we set up a regression model to compare its frequency between the public and private conditions. In each model, the dependent variable is the feedback construct (e.g., Solution), and independent variables include Comment Segment, Condition, and the Interaction between the two, as we anticipated the condition might have different effects in the three comment segments due to their different focus. We also introduced control variables indicating the quality of the original assignment submission, with the understanding that it would be harder for the reviewers to point out problems or solutions if the assignment was of high quality. As we introduced in the methods section, we operationalized assignment quality into 4 constructs, Receiver Clarity, Receiver Arguments, Receiver Arguments Unsupported, and Receiver Tradeoff. We nested each of the assignment quality indicators within Comment Segment in the regression models.

From the 9 regression models, we found that feedback generated in the public condition showed significantly higher numbers of Solution statements than feedback generated in the private condition, with F (1, 186) = 4.765, p < 0.05. The effect size value computed by Cohen’s D is 0.71, suggesting a moderate to high effect. We also found that when the assignment included more tradeoff points, there were fewer solutions pointed out in the feedback, with F (3, 192) = 3.369, p < 0.05. This demonstrates that when the assignment is of higher quality, it was harder for reviewers to propose solutions. In addition, we also saw a significant difference in the frequency of Solution Localized Idea between the two conditions, with F (1, 186) = 3.977, and p < 0.05, with an advantage to the public condition. The effect size value computed by Cohen’s D is 0.71, suggesting a moderate to high practical significance. This suggests that feedback generated in the public condition elicited more localized solutions that target at ideas.

We also found when student assignments showed a higher number of unsupported arguments and a lower number of tradeoffs, there tended to be more localized solution statements about ideas pointed out in the feedback, regardless of condition. The effect of number of unsupported arguments is marginally significant, with F (3, 186) = 2.242, p = 0.085 and the effect of number of tradeoffs is significant, with F (3, 186) = 2.839, p < 0.05. This again suggests that students tend to show more specific and substantive feedback when the assignment quality is lower. We do not see a significant difference in the frequency of other constructs of feedback between the two conditions. The average and standard deviation of the frequency for each feedback construct (i.e., count of segments that include that construct) is displayed in Table 1. These per construct descriptive statistics are offered to provide an overview of the distribution of feedback constructs between conditions, not meant as a formal between-condition comparison.

Table 1: Descriptive statistics of the 9 feedback constructs by condition

<table>
<thead>
<tr>
<th>Feedback Construct</th>
<th>Private</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td>Summary</td>
<td>0.43</td>
<td>0.50</td>
</tr>
<tr>
<td>Praise</td>
<td>0.57</td>
<td>0.50</td>
</tr>
<tr>
<td>Praise localized</td>
<td>0.15</td>
<td>0.36</td>
</tr>
<tr>
<td>Problem</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Problem localized writing</td>
<td>0.06</td>
<td>0.24</td>
</tr>
<tr>
<td>Problem localized idea</td>
<td>0.4</td>
<td>0.49</td>
</tr>
<tr>
<td>Solution</td>
<td>0.35</td>
<td>0.48</td>
</tr>
<tr>
<td>Solution localized writing</td>
<td>0.07</td>
<td>0.26</td>
</tr>
<tr>
<td>Solution localized idea*</td>
<td>0.27</td>
<td>0.45</td>
</tr>
</tbody>
</table>
Hypothesis 1B: Investigating differences in perceived quality of feedback

In this section, we present results regarding the effect of our intervention on the quality of feedback indicated by student subjective rating. We built a regression model using the Subjective Rating students gave as the dependent variable and the Condition as the independent variable. We included the assignment quality indicators as control variables, operationalized by the four constructs of assignment quality discussed above. Note that the rating was assigned to the feedback as a whole rather than each comment segment separately.

In the regression model, we found that feedback generated in the public condition received a significantly higher rating than in the private condition, with F (1, 63) = 7.238 and p < 0.01. The effect size value computed by Cohen’s D is 3.06, suggesting a very high practical significance. The number of Unsupported Arguments in the assignment predicted lower rating of the feedback by the receiver with marginal significance, with F (1, 63) = 3.65, p = 0.06.

In Hypothesis 1A, we found that our intervention successfully increased the presence of solution and localized solution at an idea specificity level in student feedback. We want to further investigate whether the higher subjective rating in the public condition came from the elevated presence of these two constructs. We added the 9 feedback constructs as independent variables into the baseline regression model as described above. In the new model, we see that Condition and Unsupported Arguments still significantly predict higher feedback ratings, with F (1, 189) = 18.3, p <0.001, and F (1, 189) = 11.44, p < 0.005 respectively. Among the 9 feedback constructs, only Problem Localized Idea is marginally significant in predicting higher rating, with F (1, 189) = 2.77 and p = 0.097. We do see students in the public condition liked their feedback better, though we do not see the effect coming from the constructs of feedback that our intervention manipulates, namely Solution and Solution Localized Idea. Through the above analyses, we confirm that hypothesis (1) is generally supported. Feedback generated in a public environment shows signs of being higher quality than feedback generated in a private environment, indicated by both expert coding of feedback quality and student subjective rating.

Hypothesis 2: Investigating the effect of condition on revision of writing

In order to test our second hypothesis, we first built a regression model to compare the revision rate between the two conditions. Though we found that there was a trend that revision rate in the public condition was higher, we did not see a significant difference between the two conditions in students’ revision rate, with F (1, 63) = 0.22, p = 0.64. This suggests that the aspects of high quality feedback manipulated through elevated accountability in our intervention may not be the same ones that are most important for achieving impact on writing. In order to understand this more deeply we conducted a post-hoc analysis.

In particular, we conducted a correlational analysis to investigate what constructs of feedback students receive are associated with increases in propensity to revise. We used the binary variable indicating whether students revised or not as the dependent variable and 4 constructs of assignment quality as control variables. We included the 9 constructs of feedback quality as independent variables. We found that the presence of Problem and Solution in feedback significantly predicts whether students revise or not, with F (1, 190) = 6.13, p < 0.05, estimate of coefficient = 0.3, and F (1, 190) = 5.00, p < 0.05, estimate of coefficient = 0.3 respectively. This is consistent with prior work that shows the presence of Problem and Solution is beneficial for student implementation of the feedback. However, we also found the presence of Summary in feedback negatively predicts whether students revise or not, with F (1, 190) = 3.93, and p < 0.05, estimate of coefficient = 0.13. This is inconsistent with prior work that found summary to be helpful for implementation. From a follow-up correlation analysis, we found that the presence of Summary is negatively correlated with the presence of Problem and Solution, and positively correlated with the presence of Praise. This suggests that when students are spending effort summarizing what they’ve seen in the proposal, they are less likely to spend time pointing out specific problems and solutions in the assignments.

In addition to students’ propensity to revise, we also investigated the revision quality differences between the two conditions. In each of the regression models, we used one of the four constructs of revision quality as the dependent variable (e.g., Revision Tradeoff), and used Condition, Occurrence of Revision, and the four constructs of assignment quality as control variables. We found the public condition showed more tradeoffs in revisions than the private condition; the effect is marginally significant, with F (1, 63) = 3.159, and p = 0.08. We do not see a difference in any other revision quality constructs.

Through the above analyses, we confirm that hypothesis (2) is only partially supported at best. There is no main effect of condition on propensity to revise between the public and private conditions. There is only a trend showing students in the public condition made higher quality revisions than students in the private condition when they did revise, and the difference is only marginally significant.
Discussion and future work

Public peer review intervenes on certain dimensions of feedback

Our experiment found that the manipulation of perceived supervision through a public peer review environment increased the presence of general solution and localized solution at an idea specificity level in student feedback. On the positive side, Solution is indicated in prior work to be the most important aspects that influence feedback quality (Nelson & Schuun, 2009; Patchan, 2016). On the negative side, more work needs to be done to achieve more substantial effects on improving feedback quality. Since the intervention of public peer review environment aims at manipulating students’ accountability in feedback writing, the result might also suggest that elevating effort to offer feedback is not the same as increasing the ability to offer feedback. The public environment and feedback prompt offers increased awareness of the aspects that are desirable in high quality feedback, but more or different support may be required to address those aspects of feedback that students lack the skill to offer. The result provided in this experiment suggests that providing a public venue for feedback is one step in the right direction, but more is needed to achieve the ultimate goal of high quality feedback.

From feedback to revision

In our experiment, we observed that the public condition increased the presence of Solution and Solution Localized Idea in student feedback and also made students like their feedback better. However, it is surprising to see that increasing the prevalence of characteristics of feedback shown in earlier work to be associated with revision of writing (Nelson & Schunn, 2009; Patchan, 2016) did not show strong effects here. There are multiple explanations of this finding. First, it might be the case that it is not enough to offer more solutions, but qualities of the solutions offered are also important. It again points to the idea that raising awareness of what component needs to be included in feedback is not enough, rather future work needs to explore how to increase students’ skill at offering feedback. Second, it could be that students lacked sufficient support on appropriation of feedback in their revision process. (Wichmann et al., 2017) This suggests that it could be fruitful to provide explicit instruction and support to students on how to incorporate feedback to revise their writing.

Implications for peer feedback research

Our experiment displayed some consistent findings with prior work in different educational settings that reinforced known mechanisms on feedback offering. For example, when the assignment is of lower quality, it is easier for students to offer solutions in their feedback; when the feedback contains Problem and Solution, the feedback receivers showed higher propensity to revise. However, it is inconsistent with prior work that when the feedback contains Summary, the receiver is less likely to revise. This could be explained by the fact that when students spent more effort writing Summary in the feedback, they spent less effort writing Problem and Solution. We consider this result to be contextual to the kinds of learning environment, where students spend minimal effort on the task. We also found that student subjective rating of feedback is not correlated with expert coding. Researchers in the future should thus be cautious of choosing metrics to evaluate feedback quality.

We acknowledge the limitations of this work as a lab study. Since the factor manipulated is related to student accountability, we must acknowledge that this audience might not demonstrate the expected effect on accountability as we would see in a higher-stakes learning environment. However, the advantage of a lab study is high internal validity, especially experimenting with factors that are not easy to directly manipulate in a real course. Coetzee et al. (2015) pointed out that though participants from crowdsourcing platforms may likely have different motivations from MOOC learners, their remote individual work setting without peer contact resembles today’s MOOC setting where most students learn in isolation. Prior work (Wen et al., 2016; Wang et al., 2017) also demonstrated the potential of such high internal validity experiments in informing subsequent high external validity deployments in real MOOCs, which will be an immediate next step of our work.

Conclusion

In this study, we investigated the effectiveness and potential of a new peer review paradigm—public peer review—where students get full exposure to assignments and reviews of the class. We hypothesized that the increased accountability of the public environment would result in higher quality feedback being exchanged. We presented the results of an experimental study comparing feedback quality and revision rate between the public and private peer review conditions. The results support our hypotheses that students in the public condition provided more solutions and localized solutions about substantive ideas in the feedback they gave, and also perceived the feedback they received to be better than students in the private condition. Though we did
not observe a difference in students’ propensity to revise their assignments between the two conditions, we see a trend showing students in the public condition demonstrated higher revision quality.

References


Purdue University, College of Science writing rubrics (retrieved on July, 2017) https://www.science.purdue.edu/Current_Students/curriculum_and_degree_requirements/writing_rubric_gray.pdf


Acknowledgements
This work was funded by NSF Grants ACI 1443068, STEM+C/IIS 1546393. Thanks to all study participants for their time, and to Anhong Guo for his help and support.
The Design and Evaluation of Optimal Computerized Guidance for Invention Activities: The Invention Coach

Catherine C. Chase, Teachers College, Columbia University, chase@tc.columbia.edu
Helena Connolly, Teachers College, Columbia University, connolly2@tc.columbia.edu
Marianna Lamnina, Teachers College, Columbia University, lamnina@tc.columbia.edu
Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University, aleven@cs.cmu.edu

Abstract: A common dilemma in educational technology is designing the optimal level and type of guidance to support open-ended learning activities. We explored this question by designing a computer-based coach for Invention activities—a form of ill-structured problem-solving followed by expository instruction. The Coach was designed to elicit the core learning mechanisms of Invention by problematizing students’ solutions and mimicking naturalistic teacher guidance. This research tests both the efficacy of our designed guidance and the appropriate amount of guidance for Invention activities. In an experimental study, 205 middle schoolers worked with full, minimal, or no guidance versions of the Coach before receiving a lecture on the target knowledge (ratio structures in science). Students who received the full guidance Coach were better able to transfer their knowledge to novel domains. The work has implications for the design of guidance in open-ended learning environments.

Introduction
A key question in the design of technology-supported open-ended learning environments, is how to effectively guide students as they engage in complex, exploratory, ill-structured, and inquiry-focused activities. Many approaches have been tried (Quintana et al., 2004; de Jong & van Joolingen, 1998). A critical challenge is to provide guidance that encourages the learner to generate and construct her own ideas, without quelling the exploratory nature of the task or robbing students of essential discovery moments (Mavrikis, Gutierrez-Santos, Geraniou, & Noss, 2013; Koedinger & Aleven, 2007). We explored this question by designing and evaluating a computer-based coach to guide students through Invention activities.

Invention is an instructional method that combines ill-structured problem-solving with subsequent instruction (Schwartz & Martin, 2004). During Invention, learners attempt to invent the deep principles of a domain. In the current research, students were inventing ratio-based equations for physical science concepts (e.g. density = mass/volume, speed = distance/time). The goal of Invention activities is to prepare students to learn from later instruction. Prior work suggests that Invention creates “a time for telling,” preparing students to appreciate the “mathematical work” of equations (Schwartz & Martin, 2004). Many studies have shown that Invention and related pedagogies boost conceptual learning and transfer to novel situations (Kapur, 2008; Schwartz & Bransford, 1998; Schwartz & Martin, 2004; Schwartz, Chase, Oppezzo, & Chin, 2011).

Of course, scaling Invention activities is difficult because students require individual guidance as they invent. Classroom studies of Invention often involve a ratio of 1 instructor to every 5 students (Schwartz et al., 2011), which is not practical for widespread adoption. Thus, we began designing and implementing a computerized Invention Coach that would provide optimal, individualized guidance (Marks, Bernett, & Chase, 2016). The Invention Coach was built using Cognitive Tutor Authoring Tools (Aleven et al., 2016), which are used to build Intelligent Tutoring Systems (ITSs). However, the Invention Coach differs from typical ITSs, which tend to provide structuring scaffolds, such as correctness feedback on problem steps and next-step hints. Instead, the Invention Coach scaffolds provide less structuring and more problematizing guidance. Problematizing scaffolds highlight a facet of the student’s work that is problematic, encourage students to grapple with deep ideas, and contradict students’ erroneous solutions (Reiser, 2004). Our Coach also differs from the Invention Lab (Roll, Aleven, & Koedinger, 2010), which favors a fairly structured approach to Invention activities. The project is part of an emerging line of research in ITSs that focuses on creating adaptive support for learners in open-ended learning environments for inquiry learning (Gobert, Sao Pedro, Raziuuddin, & Baker, 2013; Poitras & Lajoie, 2014), exploratory learning (Mavrikis et al., 2013), and learning with simulations (Borek, McLaren, Karabinos, & Yaron, 2009).

In the literature on Invention and productive failure (a related pedagogy), the level of optimal guidance is disputed. For instance, Loibl & Rummel (2014a) found that guidance in the form of contrasting cases had no effect on learning outcomes. Kapur (2011) found that intermittent teacher support and brief benchmark lessons during the Invention process hindered learning compared to a no-guidance condition. In contrast, Holmes et al. (2014) found that computer-based guidance in the form of orienting and reflection prompts led to greater
learning than unguided Invention. However, results across these studies may differ either because the type of
guidance varies with each experiment or because the guidance focuses on a single learning mechanism, when
many learning processes are at play in effective Invention activities. Because our main goal was to build an
effective system that would preserve the generative and exploratory style of Invention, we chose to provide an
indirect, problematizing style of guidance that would support multiple learning mechanisms while providing
naturalistic, human-like guidance. In this paper, we describe the process and rationale behind the Invention
Coach design. We then report on a study that tests the optimal level of guidance for Invention and evaluates the
effectiveness of the Coach’s particular style of guidance in promoting learning and transfer.

An example invention activity
Figure 1A shows an example Invention activity. The goal of the task is to invent an index of “clown
crowdedness” that describes how crowded the clowns are in each bus. Students are given a few constraints that
are necessary for solving the problem: buses from the same company are equally crowded, a bigger index
number means a bus is more crowded, the method for finding the index should be the same for all buses, etc.
Though they don’t know it, students are inventing the formula for density (density = mass/volume), where
density is conceived as a measure of how crowded clowns are in different-sized buses (e.g., #clowns/#box cars).

Invention often occurs with the aid of contrasting cases that highlight important features of a problem
solution while simultaneously illustrating the invariant structure across all cases (Bransford, Franks, Vye, &
Sherwood, 1989). Many students begin the clown crowdedness task with a simple “counting” solution, where
the number of clowns in each bus represents crowdedness, but they overlook the feature of space (or bus size).
However, by contrasting the 3-compartment Crazy Clowns bus to the 6-compartment Clowns ‘r’ Us bus in
Figure 1A, students often come to realize that bus size is a critical feature of crowdedness. Both buses have the
same number of clowns but the Crazy Clowns bus is clearly more crowded. This contrast highlights the
significance of the number of bus cars, which should give students the idea that their index must account for
space. By looking across the cases, students may induce the invariant ratio structure (clowns to bus cars) that is
common to all of them. Most students begin the task with an intuitive (but vague) notion of crowdedness which
gets further differentiated and developed as they attempt multiple Inventions.

After students finish an Invention activity, they receive some other form of instruction, often a lecture
or reading on the canonical solutions and related deep structures. Many students do not generate the correct
equation during the initial Invention phase, but attempting to do so helps them notice deep domain features and
explore how they may relate in a mathematical structure. This exploration prepares students to gain a deep and
flexible understanding of the target knowledge presented in future instruction.

Our design process
To design the optimal type and amount of guidance for Invention tasks, we took a three-pronged approach.
First, we studied teachers’ naturalistic guidance of Invention to explore how teachers walk the line between
giving and withholding assistance. Second, we implemented problematizing scaffolds, which make student
solutions “problematic” without providing direct or corrective feedback. Third, we focused on the core
cognitive processes invoked by successful Invention tasks. Finally, we created two prototype versions of the
software which were improved based on extensive pilot testing.

Study of human teacher guidance and problematizing guidance
Since computer-based guidance can sometimes feel unnatural or lack the sophistication of human teaching
tactics (du Boulay & Luckin, 2001), we modeled our system on human teacher guidance. To do this, we ran a
study of experienced science teachers guiding students one-on-one through paper-based Invention activities
(Chase, Marks, Bernet, Bradley, & Aleven, 2015). We asked teachers to guided students naturally, as they saw
fit. Gains from pretest to posttest showed that the teachers were quite successful in increasing students’
conceptual knowledge (effect size $d = 0.6$) and ability to transfer to novel domains ($d = 0.7$). Overall, we found
that teachers used an “ask more, tell less style” style of dialogue. Teachers asked questions twice as often as
they gave explanations, and they rarely gave direct right/wrong feedback. Moreover, the more teachers posed
deep questions and withheld explanations, the more students transferred their learnings to novel problems.

Given these findings, we designed the Invention Coach with an “ask more, tell less” style of guidance,
with a focus on deep questions. As such, the Coach avoids giving didactic explanations, instead prompting
students to reflect on and explain their answers. We also drew on Reiser’s (2004) construct of problematizing
scaffolds. In many educational technologies, scaffolds serve to structure, simplify, or ease the task in some way.
Problematizing scaffolds, on the other hand, add complexity to the task in the short term, by making students
confront and grapple with deep disciplinary ideas. The Invention Coach problematizes student understanding by
contradicting and poking holes in wrong solutions, and encouraging students to diagnose their own errors. Thus, the Coach avoids giving direct right/wrong feedback or explicitly stating students’ errors or how to fix them.

Supporting learning mechanisms
Our study of human teacher guidance of Invention gave us a feel for the types of prompts and dialogue to provide in the Coach, but we felt the Coach would be most effective if it also supported the key learning mechanisms of Invention (Loibl et al., 2016). There are three core learning processes invoked by Invention activities. The first is activation of prior knowledge. By asking learners to Invent their own solutions before telling them the expert solutions, we invite learners to draw out prior knowledge and skills that can then be augmented, built upon, or modified (Kapur, 2008). The second core learning process in Invention is uncovering knowledge gaps. While attempting to invent solutions and often failing, students come to realize their solution is insufficient and may identify holes in their knowledge, which they can then seek to fill during later instruction (Loibl & Rummel, 2014b). The third core learning mechanism is noticing deep features. Contrasting cases can highlight deep features of the target knowledge, preparing students to learn (from later instruction) how these features relate in a mathematical structure (Bransford et al., 1989; Schwartz et al., 2011). To maximize the effectiveness of our Invention Coach, we designed instructional modules that would support each of these three core learning mechanisms.

The Invention Coach
Drawing on our study of human teacher guidance, the problematizing framework, and the core learning mechanisms identified in the literature, we designed and implemented the Invention Coach system. Figure 1 depicts our third iteration of the Coach with an example of the crowded clowns Invention activity. In this activity students are asked to invent a numerical index to describe how crowded the clowns are in each bus. Students input their invented index numbers (B) next to each contrasting case (A). The Coach (C) provides hints and guidance along the way in the dialogue box (D). If students get stuck, they can access several resources such as a calculator, rules sheet (which describes task goals and constraints), a notepad, and a “help” button to solicit guidance (E). Students tend to invent iteratively by generating solutions, receiving guidance, then generating new solutions in Invention-guidance cycles. While the Coach never explicitly tells students whether their Inventions are right or wrong, it gives indirect feedback by posing comments, questions, or activities that keep students reflecting on and evaluating their inventions.

The Invention Coach guidance comes in several forms. Any time students “submit” their solutions to the Coach, they receive a short motivational message. We included motivational messages because we found in our pilot work that students often felt frustrated when they could not generate the right solution immediately.

Figure 1. Invention Coach main interface and forms of guidance.
(since typical school math and science problems are not solved iteratively). The motivational message is followed by either a hint or a module. Most hints remind students of problem constraints their solutions violate, while others encourage students to progress through the task (Figure 1F). Modules are longer, interactive sequences where students respond to prompts and complete activities designed to engage the learning mechanisms discussed above.

There were three main modules: ranking, tell-me-how, and feature-contrast. The ranking module was designed to help students activate their intuitive, prior knowledge of crowdedness (Kapur, 2008). In this module, learners are asked to order the companies from most to least crowded. Most students can visually distinguish between the most and least crowded bus companies (though they often do not know how to quantify crowdedness yet). This ranking activity can be a good form of early guidance for students who are initially lost. Later on in the task, students can evaluate their index numbers by comparing their intuitive ranking of the cases to the ranking provided by their indices. Some form of this ranking activity occurred fairly frequently in our study of human teacher guidance, as well. The tell-me-how module asks learners to reflect on and explain how they generated their index numbers (Figure 1G). It also mimics a question that was frequently posed by the teachers in our study of naturalistic guidance: “Tell me how you got that number.” In the tell-me-how module, students first describe their solution method in an open text box before selecting from a set of explicit strategy types derived from our piloting (“I counted”, “I estimated”, or “I calculated”). They are then pushed to describe their answers in disciplinary terms, such as mathematical expressions and units. In essence, the tell-me-how module elicits students’ mathematical self-explanations. Self-explaining is one way to surface gaps in knowledge, a key learning process in Invention (Chi, De Leeuw, Chiu, & LaVancher, 1994; Loibl & Rummel, 2014b; Roll et al., 2010). Finally, the feature-contrast module (Fig. 1G), encourages students to compare specific sets of contrasting cases to help them notice key features of the domain, another key learning mechanism of Invention (Roll et al., 2010; Schwartz et al., 2011). In the example in Figure 1H, the Coach compares two cases and asks the student why the top bus is more crowded. The comparison is designed to highlight the often-overlooked feature of bus size, as only bus size differs across the cases.

Both modules and hints are designed to enact the “ask more, tell less” style, while problematizing students’ solutions. The Coach avoids giving didactic explanations and never gives direct feedback on students’ Inventions. Rather, it points out how solutions are problematic by noting a constraint the solution violates (in hints), encouraging articulation of ideas (tell-me-how module), helping students notice a feature missing in their solution (feature contrast module), or encouraging comparison of intuitive notions of crowdedness to invented solutions (ranking module). Moreover, the majority of hints and within-module prompts are structured as deep questions, asking students to generate, explain, and connect their ideas, modeled on the teacher guidance observed in our study.

How does the Coach decide when to give which forms of guidance? Based on data from a study with a Wizard-of-Oz version of our system (where guidance was selected by a human “Wizard”), we constructed a novel model of an adaptive coaching strategy to determine when to give each kind of guidance (Aleven et al., 2017). Every time a student submits their answer, the system classifies their solution into one of five broad categories: unclassifiable, single-feature, two-feature, mathematical two-feature, and ratio. The system responds with guidance related to the most significant knowledge gap expressed in the solution. For instance, if the student’s solution only considers a single feature (e.g., number of clowns), then guidance will focus on getting the student to notice the importance of space. If the student’s solution considers both features but does not relate them mathematically, then the guidance focuses on the importance of precision (e.g., index numbers should be exact). Each solution category has a sequence of alternating hints and modules. When students submit a solution of the same type, the system provides the next piece of guidance in the sequence. Students often generate a number of different solution types in a session, jumping from one branch of guidance to another throughout.

The current study
The current experiment was designed to address two questions: (1) Can a computerized Coach that provokes key learning mechanisms via problematizing guidance effectively support learning and transfer from Invention?, and (2) How does the level of guidance during Invention (full, minimal, or none) impact learning and transfer? To answer these questions, we compared our full version of the Invention Coach system to “minimal guidance” and “no guidance” versions of the same system. This allowed us to test the efficacy of our designed guidance, while controlling for use of the computer, task structure, presence of the Coach avatar, and so on. The “full guidance” condition had access to our complete Invention Coach with hints and modules as guidance. The “minimal guidance” condition received only hint-based guidance, but no modules. This amounted to giving students guidance about the goals and constraints of the task. Thus, the “minimal guidance” condition served as a useful comparison condition when testing the Coach’s efficacy, because it does not provide additional
information that go beyond the given task goals and constraints (it merely reiterates them in different words and gives them adaptively, in response to a student’s specific solution). Finally, the “no guidance” condition received only repeated suggestions to keep working (e.g. “Keep going, you’re not quite there yet.”). Since the “full guidance” condition would have the benefit of our modules, which were designed to provoke the key learning mechanisms of Invention, we hypothesized that they would perform best on learning and transfer measures. Further, we hypothesized that the “minimal guidance” condition’s adaptive hints that problematize students’ solutions by pointing to constraints their solutions violate would produce a middle level of learning and transfer. In contrast, we predicted that the “no guidance” condition, which would have no hints or modules to help them understand why their solutions failed, would have the lowest learning and transfer outcomes. Recall that Invention activities are followed by expository instruction, so we really tested how well these various levels of guidance prepared students to learn and transfer from a later lecture.

Methods

Participants

Students from 9 seventh- and eighth-grade classes (N = 205) in a public middle school in New Jersey participated in this study over a total of five days during their regular science classes. The school population was 96% Hispanic and 56% male, with 87% receiving free and reduced-price lunch. Condition was randomized at the student level, so students within the same class were assigned to different conditions.

Design and procedure

Two weeks prior to the start of instruction, students took a pretest to assess their prior knowledge of ratio structures as they relate to density and speed concepts. Instruction lasted for three days, 35 minutes per day. On the first day, students worked with the Invention Coach on the clown crowdedness Invention activity (see Figure 1A). On the second day, students worked with the Invention Coach on a car fastness Invention activity, where their goal was to develop an index of car fastness (developing the equation for speed). Invention activities were introduced by a short video explaining task goals and constraints. Students worked individually on each Invention task. On the third day of instruction, all students received a PowerPoint lecture from an experimenter that (1) gave the scientifically accurate solution to the Invention activities (2) related these activities to the science concepts of density and speed and (3) highlighted the importance of ratio structures in physical science equations. The lecture was integrated with a set of word problems which students completed on a paper worksheet, where they practiced applying these equations in simple, well-defined problem scenarios, similar to those that would be found at the end of a textbook. The goal of this instruction was to enhance students’ understanding and ability to notice ratio structures in science. The day after this, all students completed a learning and transfer posttest.

Conditions were implemented using the full, minimal, and no guidance versions of the Invention Coach system described earlier. Each time students “submit” their indices, they received some form of guidance, which varied by condition. Additionally, all three conditions received motivational messages (e.g. “I can see the gears turning in your brain, you’re working hard!”). Also, all students had access to student-initiated tools such as a calculator, a notepad, and rules list. Both full and minimally guided versions of the system used a similar adaptive coaching strategy to select the appropriate guidance in response to a student’s solution type.

Measures

Log data from all interactions in the Coach were collected. To get a feel for the Invention problem-solving process, for each student, we calculated the number of different sets of indices submitted to the Coach, the number of unique solution types submitted (according to the five categories to which the system responds), and whether or not the student was ever able to invent a ratio-based solution.

Paper test items targeted students’ understanding of ratio structures in physical science concepts. The posttest contained 8 items: 3 conceptual, 2 application, and 3 transfer items. Conceptual items required students to reason about the ratio structure of density and speed (e.g. determining whether a large or small pillow is more “tightly packed” if the number of feathers is constant). Application items asked students to reason about density and speed ratios in novel ways (e.g. design a flower pot that has a specific density of flowers). Transfer questions assessed whether students could notice and implement ratio structures in novel domains (e.g. describe the “spray strength” [i.e. pressure=force/area] of several fountains). The pretest contained a total of 6 items consisting of a limited set of 4 isomorphic versions of the posttest items (counterbalanced), plus 2 items to test students’ prior knowledge of the density and speed equations.
Results
While working in the Invention Coach software, students submitted an average of 17 sets of indices per problem to the Coach, which did not vary significantly by condition, $F(2, 199) = 1.07, p = .35$. During this iterative process, students in the full guidance condition invented significantly fewer solution types ($M = 1.7, SE = .11$) than in the minimal ($M = 2.1, SE = .11), $p = .007$ and unguided conditions ($M = 2.2, SE = .11), p = .001$. However, the number of students who invented a ratio solution on either of the Invention activities did not differ by condition (full = 49%; minimal = 59%, none = 65%), $x^2(2) = 3.64, p = .16$.

Conditions did not differ significantly on pretest scores, $F(2, 202) = .05, p = .96$. To analyze how condition affected posttest measures, we conducted a MANCOVA, with the three posttest measures (conceptual understanding, application, and transfer) as dependent variables, and condition, class, and whether students invented a ratio solution as independent variables, and pretest score as covariate. All variables had significant main effects on the dependent variables ($p$’s < .02). Interactions were not significant, so they were excluded from our MANCOVA model. The omnibus MANCOVA revealed a significant main effect of condition, $F(6, 382) = 2.50, p = .02, \eta^2 = .04$. Follow-up ANOVAs revealed that all three groups performed similarly on conceptual and application items, $p$’s > .14, but differed significantly on transfer items, $F(2, 192) = 5.04, p = .007$. As shown in Table 1, the descriptive pattern on posttest transfer items is that the full guidance condition performed the best, followed by the minimal guidance condition, and then the no guidance condition. However, posthoc tests (with Bonferonni correction) on posttest transfer scores revealed that only the full guidance condition performed significantly better than the no guidance condition, $p = .005$, however the minimal guidance condition did not differ significantly from either condition.

Table 1. Adjusted means and standard errors for each item type on posttest. Max scores = 1

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Full Guidance</th>
<th>Minimal Guidance</th>
<th>No Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
<td>M</td>
</tr>
<tr>
<td>conceptual</td>
<td>0.71</td>
<td>0.02</td>
<td>0.68</td>
</tr>
<tr>
<td>application</td>
<td>0.49</td>
<td>0.04</td>
<td>0.56</td>
</tr>
<tr>
<td>transfer**</td>
<td>0.55</td>
<td>0.03</td>
<td>0.50</td>
</tr>
</tbody>
</table>

To compare the efficacy of our full Invention Coach to the human teachers in our previous study (Chase et al., 2015), we explored gains on items that were common on pre and posttests. While there were no significant gains on conceptual items, the full guidance group made significant gains on the common transfer item, effect size $d = 0.6$. Thus, the full guidance provided by the Invention Coach was almost as effective as the human teachers used in our previous study in enhancing transfer ($d = 0.7$), but not conceptual learning ($d = 0.6$). While this is not the most rigorous comparison given differences in the participant populations and time spans across these two studies, it adds evidence to the conclusion that our system is effective.

Discussion
We have detailed the design of the Invention Coach, which was based on a study of human teacher guidance of Invention along with prior research and theory on the cognitive processes that make Invention effective at enhancing transfer. Our full guidance version of the Coach has an adaptive, “ask more, tell less” style of scaffolding students through Invention activities, when the goal is transfer. Our results indicate that in computer-based learning environments, scaffolding Invention tasks can be effective, when the guidance substantially problematizes students’ solution attempts and encourages core learning mechanisms.

This study also demonstrated that extensive guidance provided by the full guidance version of the Coach did not hinder student exploration. While students in the full guidance condition created fewer solution...
types, they submitted the same number of solutions as the other conditions, and students in the full guidance were just as likely to generate the correct solution as students who received minimal or no guidance.

However, we also found that the full guidance Coach was not effective at enhancing conceptual learning or the ability to apply and manipulate learned ratios. There are two explanations for this outcome. One is that students may have learned these concepts and applicative abilities largely from the lecture and practice that followed – which were identical for all three conditions. Unfortunately, we cannot present hard evidence of this, since the worksheets students completed during the lecture were not done independently. A second possibility is that Invention pedagogies are uniquely designed to enhance transfer, but not learning, as has been found in other work (Schwartz & Martin, 2004). Thus, guidance designed to maximize the benefits of Invention may improve transfer only.

A limitation of this work is that the current Coach is only equipped to support Invention of ratio-based equations. In future work, we aim to adapt the Coach to support Invention of a broader variety of equation types (additive, multiplicative, exponential, etc.). A second limitation is that we have not isolated exactly what makes the Invention Coach effective in comparison to the “no guidance” condition. Future work could test the efficacy of problematizing guidance and the addition of guidance modules separately.

Despite limitations, this research makes both practical and theoretical contributions. The work adds to the small body of evidence regarding the question of how much guidance might be optimal during Invention activities. The results seem consistent with that of Holmes et al. (2014), who found that a modest amount of scaffolding led to better conceptual learning outcomes. Our results diverge from Loibl and Rummel (2014a) and Kapur (2011), who found that guidance during early instruction was either detrimental or simply not helpful. In the case of Loibl & Rummel (2014a), the lack of effect may be due to their focus on a singular learning mechanism, while guidance may need to address the full suite of learning processes to be effective. In contrast, our guidance (as well as that of Holmes et al., 2014) taps into additional, related learning processes such as identifying knowledge gaps and activating prior knowledge. In the case of Kapur (2011), it is possible that the guidance, some of which contained mini lectures, provided too much structure, dampening the exploratory nature of the task (Chase et al., 2015). Thus, future research could investigate not just how much guidance is optimal in Invention, but also what kind of guidance produces the greatest learning gains.

More broadly, we have created the first technology designed to support Invention with adaptive guidance, modeled on human teacher guidance. This adds to the body of work on the efficacy of various forms of technology-based scaffolds for open-ended learning tasks (Quintana et al., 2004; de Jong & van Joolingen, 1998; Reiser, 2004). Moreover, this paper provides a blueprint for designing software to guide open-ended problem-solving with deep questions, few explanations, problematizing guidance, and modules that invoke specific learning mechanisms.

References


**Acknowledgments**

We are grateful to Jenna Marks and Octav Popescu for their essential contributions to the Coach’s design and implementation. This work was supported by the National Science Foundation under Grant No. 1361062.
Matching Data-Driven Models of Group Interactions to Video Analysis of Collaborative Problem Solving on Tablet Computers

Luc Paquette, Nigel Bosch, Emma Mercier, Jiyoon Jung, Saadeddine Shehab, and Yurui Tong
lpq@illinois.edu, pnb@illinois.edu, mercier@illinois.edu, jiyoonj@illinois.edu, shehab2@illinois.edu, ytong13@illinois.edu
University of Illinois at Urbana-Champaign

Abstract: Despite an increasing emphasis on the use of collaborative learning in classrooms, there is still much to be understood about how to successfully implement it. In particular, it is still unclear what the role of teachers should be during collaborative learning activities and how we can better support and guide teachers in their implementation of collaborative activities. In this study, we investigated how digital learning environments can be leveraged to support collaborative learning through data-driven models of students’ collaborative interactions by matching video and log data. The models successfully detected off-task behavior (43.2% above chance-level accuracy) and task-related talk (34.5% above chance) as students solved problems using a collaborative sketching tool. Future work will investigate how these models can be used to allow instructors to intervene effectively to support collaborative learning through the use of data-driven tools which will provide them with live information about the students’ behaviors.

Major issues and theoretical approaches
Collaborative problem solving is an important skill (Hesse, Care, Buder, Sassenberg, & Griffin, 2015), and its prominence in international education and assessment systems has been increasing (e.g. ABET, 2015; NRC, 2012; OECD, 2017). However, there is still much to be understood about how to successfully implement collaborative learning in classrooms (Nokes-Malach, Richey, & Gadgil, 2015), and in particular, how teachers can be most effective in supporting students’ interactions (e.g. Webb et al., 2009). Kaendler and colleagues (2015) identified a key role that teachers play in monitoring and intervening when groups struggle, and prior work indicates the relevance of teacher interventions is important for successful group outcomes (e.g. Dekker & Elshout-Mohr, 2004). However, while master teachers are more likely to have developed expertise about assessing how and when to intervene, more novice teachers may struggle with this. For example, earlier work has shown that almost all interventions made by graduate teaching assistants were content focused, with very few interventions focused on supporting students’ collaborative interactions (Mercier, Shehab & Kessler, under review). Thus, there is a need to explore ways to provide insight into the group processes for novice teachers, allowing them to understand more about what is going on within groups, and intervene appropriately (e.g. Alavi & Dillenbourg, 2012).

In this paper, we present initial work towards creating a data-driven teacher tool that automatically provides such insight. Building on a project that sought to create a shared representation tool for engineering students, we developed prediction models by matching log data from the student tool to video analysis of their interaction behaviors. Our results indicate that there is potential in using logs of student actions to assess the quality of their interactions, which could be implemented in a teacher tool to augment their observations and provide insight into when and how best to intervene.

Collaborative learning
The value of collaborative learning for both learning and transfer and as a way to increase persistence and interest in STEM fields has been identified across a range of studies (e.g. Barron 2003; Gasiewski et al., 2012). However, variation in the quality of outcome has been recorded in both classroom and laboratory studies (Nokes-Malach et al., 2015), and success is most often associated with the quality of student interactions during collaborations (Kaendler, et al., 2015). Hesse and colleagues (2015) identified behaviors that are most associated with successful interactions, dividing them into social skills and cognitive skills. There is an increasing recognition that students need help developing these skills, and that merely placing students in groups is not sufficient for groups to function well (Authors, under review; Borge & White, 2016). Teachers play a key role in intervening to support groups as they develop these skills, needing to make a quick assessment as to whether students need support in relation to the social or cognitive collaborative processes, or in relation to the course content. Initial work in this area points towards productive insights being provided either by student actions—for example, by changing the color of a lamp (Alavi & Dillenbourg, 2012) or posting a tweet (Mercier, Rattray, & Lavery, 2015)—or by insight automatically provided to the teacher by the software the students are using (Martinez-Maldonado, Yacef, & Kay,
Thus, our primary question in this paper is how can we automatically detect students’ collaborative interaction patterns and use them to provide insight to teachers?

Modeling student behavior from action logs
Researchers in the field of Educational Data Mining (Baker & Yacef, 2009) have studied how machine learning approaches can be used to build student models that are able to detect when students using digital learning environments are engaging in specific behaviors, or to infer the student’s current state of mind. This is achieved by collecting detailed logs of students’ actions within the learning environment. Those logs are then analyzed using machine learning algorithms to find relationships between the students’ actions and the modeled construct. For example, a model might learn that repeatedly submitting the same answer on a homework problem is indicative of frustration.

Action logs have been used to model a variety of constructs across multiple types of digital learning environments, such as students’ disengagement in intelligent tutoring systems, where models were trained to detect when students attempt to “game the system” (Baker, Corbett, Roll, & Koedinger, 2008; Paquette, de Carvalho, Baker, & Ocumpaugh, 2014). Gaming the system is a type of off-task behavior in which students exploit a computerized tutor’s functionalities to guess an answer or have the tutor provide them with the answer. Sabourin, Mott, and Lester (2013) studied how action logs can be used to model self-regulated learning behaviors in an educational game called Crystal Island. Gobert, Sao Pedro, Raziuddin, and Baker (2013) used action logs to assess whether students were showing behaviors related to the usage of science inquiry skills. In addition, action logs have been used to detect students’ states of mind, such as their affective states (Baker et al., 2012; Kai et al., 2015; Paquette, et al., 2014; Pardos, Baker, San Pedro, Gowda, & Gowda, 2014) or whether they are mind-wandering (Mills & D’Mello, 2015). This type of research has been conducted in many types of learning environments including intelligent tutors, educational games, and science simulation microworlds.

C-STEPs
The study presented in this paper was conducted using C-STEPs (Collaborative Support Tools for Engineering Problem Solving). This software is a web-based application whose main functionality is to provide students with a shared digital environment to support the creation of joint representations while engaged in collaborative problem solving. C-STEPs is presented on tablets, which are synchronized for each group, so members of a group share their work (and therefore, their problem-solving representations) with their teammates (Figure 1). As the students use C-STEPs, a detailed record of their actions is stored in logfiles.

In this paper, we focus on connecting students’ actions within C-STEPs to video analysis of their collaborative interactions between group members. The specific research questions addressed in this paper are:

1. Are there associations between the types of interactions identified in the video and the patterns in the logfile data?
2. What insight into group processes do the logfile data provide that could inform teachers about the status of the group or appropriate intervention strategies?

Methods
Design
This study is part of a multi-year design-based implementation research project focused on collaborative learning in a large introductory engineering course. This project aligns with college goals to increase the use of collaborative learning across core introductory courses, and students engage in collaborative problem-solving activities in all discussion sections in this sequence. The research team work closely with the faculty (some who
are on the research team) and TAs on task design and cultural change issues within the program. Four discussion sections (classes) attended their regularly scheduled class in an instrumented lab classroom for three consecutive weeks during which the data for this study was collected. During those sessions, students were grouped in teams of four (although due to attendance issues, group sizes ranged from 2-5) to complete collaborative engineering tasks using C-STEPS on either a set of tablet computers or a multi-touch table.

This paper focuses on students who used the tablet computers while solving the tasks. Multi-touch table data were not included because logged actions were fundamentally different. Data from week 1 were also discarded due to data collection issues, and as this week was seen as an introductory week for students to become familiar with the software. In total, 82 unique participants (25 female and 57 male) used the tablet computers in 14 groups in week 2 and 11 groups in week 3. All of these students gave consent to participate in the study.

Data sources
We used data collected from two sources. First, video and audio recordings were collected, providing us with rich information about how students in each group collaborated with each other. Second, action logfiles, containing a detailed list of all the students’ actions on the tablets, were collected. Each logfile contained information necessary to reconstruct the students’ problem-solving process. This included lists of point coordinates for each of the students’ drawing segments, records of when students used clear screen and undo functionalities, screen scrolling actions, and other actions. In addition, timestamps were recorded to indicate the exact time of each action.

Data from videos and logfiles were synchronized with each other, allowing us to associate the students’ actions on the tablets (logfiles) to their collaborative interactions (observed through video). The synchronized data were then segmented in one-minute clips for further analysis. We chose a length of one minute so that clips were long enough to observe meaningful collaboration behaviors while simultaneously being short enough to reduce the chance of observing multiple collaborative interactions. In total, 1,128 clips were extracted.

Coding of collaborative interaction
A subset of the videos was used to develop an emergent coding scheme of types of student interactions. The codes are shown in Table 1. The task relatedness, peer interaction, and tablet usage dimensions involved codes that were mutually exclusive; e.g., a group was either on task or off task during a one-minute clip. The talk content and teaching assistant (TA) interaction dimension were not mutually exclusive and each specific code was marked as being observed if a significant portion of the one-minute clip involved the code. For example, a clip could be marked as containing both task-related talk and other talk.

To determine reliability, two trained coders independently coded 60 video clips. The interrater reliability (Cohen’s kappa) is reported in Table 1. Once interrater reliability was established, the remaining clips were coded by the two coders. Table 1 provides a count of how many instances of each code were identified in the dataset.

These codes were selected for this initial work, with a recognition of the need to expand the coding scheme to more complex collaborative behavior in later work. They are also drawn from prior research with this population, where there is evidence of TAs disrupting productive on-task collaborations when they intervene. One goal of the interface is to help TAs recognize good collaborations, as well as intervene to support groups.

Logfile feature extraction
Logfiles were processed to compute features that indicated how groups of students used the tablets during each one-minute clip. The features we computed provide information about 1) the total quantity of actions, such as how many line segments were drawn, how many times the screen was cleared, or how many time students undid their last action; 2) the location of the actions, such as the horizontal and vertical positions of drawings; and 3) student co-interaction, such as how many different students drew on the tablet during the clip and the difference between the students’ scroll position. Overall, 28 features were computed from the logfile. A detailed list of those 28 features is provided in the result section (Table 3).

Detecting collaborative behaviors from logfile data
Predictive models were trained to detect the students’ collaborative interactions using RapidMiner 5.3 (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006), a graphical data mining toolkit that allows users to apply machine learning algorithms to their data without requiring computer programming expertise. Each model was designed to detect whether students were engaged in a specific type of interaction or not. For example, whether students were off-task or not, engaging in task relevant discussion or not, and so on. Models were created for codes related to off-task behavior, no talk, task-related talk, other talk and peer interaction (Table 1). We did not create models for TA interaction, since this information would not be relevant to report to the TA, and tablet usage, since the action logs already provided us with detailed information about the groups’ usage of the tablets.
Table 1: Coding scheme of student interactions, including interrater reliability and number of each code identified

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Code</th>
<th>Definition and Examples</th>
<th>Cohen's kappa</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task relatedness</td>
<td>On task</td>
<td>The clip shows that at least one of the group members appears to be on task (e.g. two students solving problems on the tablet)</td>
<td>0.95</td>
<td>905</td>
</tr>
<tr>
<td></td>
<td>Off task</td>
<td>The clip shows that all group members appear to be off task (e.g. two students are texting using their phone)</td>
<td></td>
<td>223</td>
</tr>
<tr>
<td>Talk content</td>
<td>No talk</td>
<td>The clip shows that group members are not talking to one another or to the teaching assistant (TA)</td>
<td>0.88</td>
<td>207</td>
</tr>
<tr>
<td></td>
<td>Task talk</td>
<td>The clip shows that at least two group members or at least one group member and the TA are talking about task-related topics</td>
<td></td>
<td>707</td>
</tr>
<tr>
<td></td>
<td>Other talk</td>
<td>The clip shows that at least two group members or at least one group member and the TA are talking about topics that are not related to the task (e.g. socializing or technology issues)</td>
<td></td>
<td>278</td>
</tr>
<tr>
<td>Peer interaction</td>
<td>No peer interaction</td>
<td>The clip does not show any verbal interaction between group members</td>
<td>0.87</td>
<td>357</td>
</tr>
<tr>
<td></td>
<td>Peer interaction</td>
<td>The clip shows verbal interactions between group members</td>
<td></td>
<td>771</td>
</tr>
<tr>
<td>TA interaction</td>
<td>Not present</td>
<td>The clip does not show any presence of the TA</td>
<td>0.91</td>
<td>682</td>
</tr>
<tr>
<td></td>
<td>Whole class</td>
<td>The clip shows that the TA is interacting with the whole class (e.g. the TA is making a whole class announcement)</td>
<td></td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>The clip shows that the TA is interacting with at least one group member</td>
<td></td>
<td>316</td>
</tr>
<tr>
<td>Tablet usage</td>
<td>Tablet used</td>
<td>The clip shows that at least one of the group members had their eyes, fingers, or stylus on the tablet</td>
<td>1.00</td>
<td>1034</td>
</tr>
<tr>
<td></td>
<td>Not used</td>
<td>The clip shows that none of the group members were using the tablet</td>
<td></td>
<td>94</td>
</tr>
</tbody>
</table>

Model training
Since the type of relationship between the students’ collaborative interactions and their actions on the tablet was unknown (e.g., linear or piecewise relationships), we initially tested three algorithms that capture different types of relationships: C4.5, a decision tree based approach; RIPPER, a decision rule based approach; and naïve Bayes, a probability distribution based approach. Those well-known algorithms were chosen based on related research in digital learning environments (Baker et al., 2012; Pardos et al., 2014). The naïve Bayes algorithm matched student actions best, and was selected for our final models.

Model performance evaluation
We evaluated models with two different performance indicators: Cohen’s kappa (Cohen, 1960) and AUC (Area Under the receiving operating characteristic Curve) computed using the A’ approach (Hanley & McNeil, 1982). Kappa is computed in the same way as when assessing interrater agreement and measures the degree to which a model is better than chance at identifying a group’s behavior. In this context, a kappa of 0 indicates a model that performs at chance level whereas a kappa of 1 indicates a perfect model. Kappa is useful for evaluating models with imbalanced codes (e.g., 80.23% on-task and 19.77% off-task behavior) because it adjusts chance level for imbalance, unlike accuracy, which may be skewed by predicting the most frequently observed interaction (e.g., 80.23% accuracy by predicting everything as on-task given its prior proportion in Table 2).

AUC was computed using the A’ approach (Hanley & McNeil, 1982). A’ is the probability that, given two examples of different codes, the model will be able to correctly classify the examples. Thus, an A’ of 0.5 indicates a model that performs at chance level, whereas an A’ of 1 indicates a perfect model. Unlike kappa, A’ evaluates the model across all possible tradeoffs between correct predictions of the positive code (the code of interest) and incorrect predictions of the negative code. This provides a complementary perspective on model performance in conjunction with kappa.

Model validation
Each model was validated using five-fold group-level cross validation. Using this approach, the full set of 1,128 clips was randomly separated into 5 subsets, each including the data for 5 of the 25 groups. Then, predictive models were trained using 4 of the 5 subsets with the remaining subset used as a held-out test set. This process was repeated five times so that each of the 5 subsets was used as the held-out test set exactly once. Performance of the model was then evaluated on the aggregated predictions of the five held-out test sets. By going through this process, we evaluated the performance of the models when predicting student interactions for new (unseen) groups of students. This was done to avoid reporting results of models that were over-fit to the training data.

Feature selection
Forward selection was used during model training, within each cross-validation fold, to find the smallest set of features that produces the best predictive model. Using this approach, features were introduced in the model one at a time, based on their predictive power. First, predictive models including only one feature were trained with each of the available features. The feature resulting in the model with the highest performance (measured as the sum of kappa and $A'$) was added to the set of selected features. Then, additional models were trained combining the selected feature and each of the remaining features, producing a list of models built using two features. Out of those, the best one was selected and the newly added feature was included in the set of selected features. This process was repeated, adding one new feature each time, until no increase in performance was observed.

Results
Table 2 provides a summary of the performance of the trained predictive models for each of the five predicted types of interactions. Proportions of the behaviors are reported to provide information about data imbalances. For example, in our data, 19.77% of clips were coded as off task (223 out of 1,128 clips). Note that the proportions do not sum to 100% as none of the five behaviors codes were mutually exclusive.

Table 2: Performance metrics and number of selected features for each of the five predictive models

<table>
<thead>
<tr>
<th>Type of Collaborative Interaction</th>
<th>Proportion</th>
<th>Kappa</th>
<th>$A'$</th>
<th># Selected Logfile Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off task</td>
<td>19.77%</td>
<td>0.432</td>
<td>0.748</td>
<td>10</td>
</tr>
<tr>
<td>No talk</td>
<td>18.35%</td>
<td>0.231</td>
<td>0.650</td>
<td>8</td>
</tr>
<tr>
<td>Task talk</td>
<td>62.68%</td>
<td>0.345</td>
<td>0.683</td>
<td>7</td>
</tr>
<tr>
<td>Other talk</td>
<td>24.65%</td>
<td>0.135</td>
<td>0.541</td>
<td>14</td>
</tr>
<tr>
<td>Peer interaction</td>
<td>68.35%</td>
<td>0.225</td>
<td>0.682</td>
<td>4</td>
</tr>
</tbody>
</table>

We investigated which of the students’ actions on the tablets were predictive of collaborative interactions by examining individual features that were selected by each predictive model. For each feature, the naïve Bayes algorithm fits two normal distributions, one for positive prediction (e.g., the students are talking about the task) and one for the negative prediction (e.g., the students are not talking about the task). This results in pairs of means and standard deviations associated to each of the logfile features in the model. We used those values to calculate effect sizes, using Cohen’s $d$, that show how much the logfile features differed between predicted codes. Table 3 provides a complete list of the 28 features used in the predictive models, as well as $d$ for each selected feature (blank spaces indicate that the feature was not selected).

Conclusions and implications
Our findings show that students’ action on a collaborative tool on tablet computers were indicative of their collaborative interactions with each other. As can be seen in Table 2, model performance was uneven, ranging from kappa = 0.135 to 0.432 and $A' = 0.541$ to 0.748, but each model performed above chance level. We expect that the two predictive models that were most successful, off task (kappa = 0.432; $A' = 0.748$) and task talk (kappa = 0.345; $A' = 0.683$), will be particularly useful for informing teachers about the status of the groups as they solve the task. Indeed, although it can be easy for a teacher to observe whether students are touching their tablets; it is difficult for them to quickly evaluate whether student actions are on task without focusing their attention on each individual group. Similarly, it can be difficult for teachers to evaluate whether the students’ discussions are related to the task without making an effort to listen to the content of conversations.

Further analysis of the features selected for each of our predictive models shows which of the students’ actions and patterns of actions were most predictive of the types of interactions the students engaged in. As can be seen in Table 3, some features, such as the cumulative number of events for a group and the number of students who executed at least one action on their tablet within the one-minute clip, were selected for multiple models and
had larger differences between behaviors (as measured using Cohen’s $d$). Conversely, other features, such as the minimum and maximum horizontal positions of drawings, were infrequently selected or had small effect sizes.

### Table 3: Cohen’s $d$ (absolute value) for each of the 28 logfile features used to build predictive models

<table>
<thead>
<tr>
<th>Logfile feature</th>
<th>Off-Task</th>
<th>No-Talk</th>
<th>Task-Solving</th>
<th>Other Talk</th>
<th>Peer Inter.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of total events</td>
<td>0.169</td>
<td>0.062</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative number of events in the session</td>
<td>0.707</td>
<td>0.434</td>
<td>0.331</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Proportion of time students spent drawing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.532</td>
</tr>
<tr>
<td>Number of lines drawn</td>
<td>0.185</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total length of lines drawn</td>
<td>0.207</td>
<td>0.323</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of points drawn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of points erased</td>
<td>0.106</td>
<td>0.179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of times students used pointer functionality (displays a temporary dot on all tablets in the group)</td>
<td>0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of times entire screen was cleared (all drawing erased)</td>
<td>0.111</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of undo actions</td>
<td></td>
<td>0.202</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of students who drew at least once</td>
<td></td>
<td></td>
<td>0.704</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of students who scrolled at least once</td>
<td></td>
<td></td>
<td>0.078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of students who performed at least one action on their tablet</td>
<td>0.519</td>
<td>0.627</td>
<td>0.447</td>
<td>0.608</td>
<td></td>
</tr>
<tr>
<td>Proportion of time students spent executing actions on the tablet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum number of students simultaneously executing actions on the tablet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.161</td>
</tr>
<tr>
<td>Number of times students scrolled</td>
<td>0.007</td>
<td>0.025</td>
<td>0.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation of scroll position while scrolling (captures speed)</td>
<td>0.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of scroll positions</td>
<td>0.047</td>
<td>0.127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum difference between different students' scroll positions</td>
<td>0.547</td>
<td>0.250</td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean difference between different students' scroll positions</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum horizontal position of drawings</td>
<td></td>
<td></td>
<td></td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>Maximum horizontal position of drawings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum vertical position of drawings</td>
<td>0.144</td>
<td>0.181</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum vertical position of drawings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal position of the center of mass of drawings</td>
<td>0.460</td>
<td>0.221</td>
<td>0.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical position of the center of mass of drawings</td>
<td>0.445</td>
<td>0.032</td>
<td>0.309</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal position of the drawing center of mass relative to the horizontal range of drawing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical position of the drawing center of mass relative to the vertical range of drawing</td>
<td>0.037</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As expected, the number of students who performed at least one action on their tablet during the one-minute clip, was a strong predictor of the types of collaborative interactions. This feature was selected for four of the five models and achieved some of the highest values for $d$ (ranging from 0.447 to 0.627). This strong predictive power is interesting as it suggests that the students’ collaboration is indeed observable in their tablet activity, rather than the collaboration only being observable outside of the collaborative tool.

The cumulative number of interaction events since the beginning of the task was also a strong predictor. It was one of only two features that was selected in four different models, and it had large $d$ values (up to 0.707) in most models except for “other talk” ($d = 0.006$). This finding is interesting since the cumulative number of events does not simply capture the events within the current one-minute clip. Rather, it indicates how much students used the tablets since the beginning of the task. As such, it suggests that prior behavior is predictive of current collaboration.
Both the vertical and horizontal position of drawings on the worksheet were effective predictors of the types of student interactions. Vertical position of the drawing is an indicator of how far the students have made it through the worksheet, since students tend to work from top to bottom. Similarly, writing is usually done from left to right. As such, the horizontal position of actions on the worksheet can be used to identify productive work since unproductive drawing (e.g., doodling) is less likely to start at the left side of the worksheet. Overall, the vertical positions of the drawing interactions were more predictive than the horizontal positions, and the center of mass of the drawing was more predictive than other indicators of location (e.g., as minimum and maximum positions).

The students’ scroll position were also predictive of key types of student interactions, perhaps because scroll features can be an indicator of progress on the task. Students who are scrolled further down are more likely to have made more progress towards completing the task. The strongest predictor related to scroll position was the maximum difference between students scrolling position, an indicator of whether students are paying attention to the same part of the worksheet.

This work is in its early stages and future data collection and model refinement will be necessary to improve predictions and incorporate them into tools for teachers. However, the potential of this method is clear from these initial findings. Future work will focus on improving models by taking advantage of additional data sources that are available in a live classroom setting. For example, Viswanathan and VanLehn (2017) have successfully combined audio data, collected using unidirectional headset microphones, with action logs to identify different types of collaboration. Although distributing individual headsets to each student in a live classroom is not feasible, we are investigating how the audio captured by the tablets’ integrated microphones can improve our models. Similarly, future work will explore how the tablets’ accelerometers can be used to improve models. Data from accelerometer could be used to give us insights about when students turn and move their tablets—for example, to show their perspective to a teammate. In addition, this early work identified simple interaction behaviors, while later work will address behaviors drawn from the research on successful groups.

Supporting students during collaborative learning is essential for effective implementation of this form of pedagogy, which is being used more extensively across higher education. However, we have found that instructors rarely have the expertise to assess and intervene successfully to support collaborative interactions (Mercier, Shehab & Kessler, under review), focusing almost exclusively on content rather than feedback to groups about their problem-solving processes. There is value of giving instructors real-time insight into group processes during class, rather than relying on their prior knowledge of collaboration to decide how to intervene or retroactively analyzing groups through video analysis. By matching tablet action logs to video analysis, we plan to leverage the automatic analysis of student actions on tablets to give instructors insight into which groups need attention, and perhaps even guidance about the types of intervention that might be needed. Such guidance will not only improve collaborative learning for students, but also teach instructors about collaborative interactions and help them to better assess groups themselves. Further research will address the question of how to best provide insights to the instructor in an actionable and meaningful way.

References


Acknowledgments
This work was supported by the National Science Foundation (NSF) through the NSF Cyberlearning and Future Learning Technologies Program (award numbers: 1441149 and 1628976).
Conjecture Mapping the Library: Iterative Refinements Toward Supporting Maker Learning Activities in Small Community Spaces

Victor R. Lee, Mimi Recker, and Abigail L. Phillips
victor.lee@usu.edu, mimi.recker@usu.edu, abigail.phillips@usu.edu
Instructional Technology and Learning Sciences, Utah State University

Abstract: A recent and important innovation in design-based research (DBR) is the use of conjecture maps, where researchers explicitly articulate the conjectured mediational relations between their designed goals and the learning designs and contexts. In this paper, we present a design case as an iterative sequence of evolving conjecture maps. As each conjecture map was tested, we revised it to highlight and refine our articulation of the tools and processes that embodied our design approach. Our design case involves small-town and rural community and school libraries in the United States as partners and DBR sites, with the goal of supporting librarians as they learn to develop and enact new STEM-oriented maker programs for youth. We show how conjecture mapping informed and supported our DBR work and how it helped push for specificity in hypothesized relations between the design, the learning setting, the outcomes, while also forcing a reflection on design constraints.

Introduction
A hallmark of the Learning Sciences in the decades since its inception is design-based research (DBR) (DBRC, 2003). While the digital age portends the increased availability of new technologies to support learning, we contend that design-based research, with its focus on understanding and refining practical and theoretical knowledge about how change is effected in real world use contexts, remains both important and necessary to the Learning Sciences scholarly community. To that assertion, however, we note that efforts to critically reflect upon and modify the work and practices of DBR have been fairly limited. Relatively recently, there have been new proposals for modifying the stances and routines of DBR, such as social design experiments – where equity and issues of justice are top-level foci (Gutiérrez & Jurow, 2016), design-based implementation research – where large educational systems are enlisted as sites and partners for designs addressing problems of practice (Penuel et al., 2011), and conjecture mapping – where design researchers explicitly specify the conjectured mediational relations between their goals and the learning environment (Sandoval, 2014). The latter is the focus of this paper. In following with a more recent turn in DBR research toward presenting design cases and how they adapt over time (Svihla & Reeve, 2016), we present our design case for inspection.

Our design case involves small town and rural public and school libraries in the United States as partners and DBR sites. Recently, libraries have been generating interest in Learning Sciences as designed learning spaces (Lee et al., 2016; Subramanian et al., 2012). Successful implementation of new learning environments and programs within libraries have typically involved large libraries in urban communities, with the YOUMedia program in Chicago (Austin et al., 2011) and the Bubbler in Madison (Halverson et al., 2017) as noted examples in the United States. Less well represented in the Learning Sciences literature are libraries that serve small towns and rural communities. Libraries serving these communities offer unique challenges: their staffing levels tend to be lower, staff are less likely to hold appropriate library certification or degree, and resources are typically insufficient (Barron, 1995).

In this paper, we begin by describing the settings for our work, and our goal of supporting librarians as they learn to develop and enact new STEM-oriented maker programs for youth. We present this work as an evolving sequence of conjecture maps that articulate theoretically important features of our design and how they were expected to produce intended outcomes. As each conjecture map was tested, we revised it to highlight and refine our articulation of the sociotechnical tools and processes that embodied our design approach. In this way, we show how conjecture mapping reflected, informed, and supported our DBR work. In following with other published models in the Learning Sciences (e.g., Edelson, Gordin, & Pea, 1999), we identify how our understanding of a complex learning setting and leverage points developed over time. We also articulate what conjecture mapping offer to learning scientists over an extended period of design research.

Literature review
The origin of design-based research is often traced to the early 1990s and the writings of Ann Brown (1992) and Allan Collins (1992). While the terminology originally differed in those early works (i.e., design experiments and design science, respectively), the fundamental orientation remained the same: researchers aimed to establish...
practical theory from creating, testing, and refining educational resources and learning designs in real world settings. The educational theory to be developed, termed by some as “humble theory” (Cobb et al., 2003), would knowingly differ from the kinds of grand theories associated with laboratory science which often attempted to uncover fundamental and unifying principles and laws that transcended contexts and setting. For education, which involved inherently complex relations and practices, an alternative research approach was needed that could support the discovery and realization of educational goals. As an example, design-based research could establish new kinds of scientific practices that could be realized in a K-12 classroom (e.g., argumentation) and articulate the means and tools necessary to get there.

Conjecture mapping (Sandoval, 2014) has been introduced as a means for such articulation. It arose from the observation that while there appeared to be uptake in the use of DBR (Anderson & Shattuck, 2012) and in its sanctioning within official publications (e.g., Sandoval & Bell, 2004), processes for doing DBR were ill-specified. At its core, conjecture mapping involves the following elements: 1) a high level conjecture about how to support learning, 2) the embodiment of that conjecture in a specific design, 3) mediating processes produced by the embodiment to yield 4) desired outcomes. The relationship between the embodiment of the design and the mediating processes are the design conjectures. The relationship between the mediating processes and the outcomes are the theoretical conjectures. Through conjecture mapping, the simultaneous work of developing a design to support learning and developing theory can both be brought to the fore. Sandoval (2014) represents these relations in a manner comparable to Figure 1.

![Figure 1. The grammar of conjecture mapping.](image)

Informally, this grammar for describing design-based research has been enthusiastically received by the Learning Sciences community, but instances of it have rarely appeared in the literature (W. Sandoval, personal communication, May 12, 2017). Throughout this paper, we use the grammar and vocabulary of conjecture mapping to present the iterative (and still evolving) conjectures we have been developing to support the following initial high-level conjecture: given appropriate supports, libraries is small and rural communities can provide a range of Maker and STEM-oriented educational programs for youth patrons that enable production of functional and digitally enhanced artifacts. Similar to how others have used high level conjectures, ours includes a vision of learning and learning environments and how particular forms of it can and should take place in library settings. As will be seen below, this high level conjecture, the embodiment of design decisions, mediating processes, and outcomes have changed over time to help us better understand the domain and the supports necessary to engineer our intended learning activities.

**Data sources**

Partners for this design work include four middle school libraries from a single, rural-serving school district in the United States and two public libraries in small towns within the same school district boundary. Data have been collected at virtually all of the libraries and with all librarians at these sites, with three school libraries and one public library providing the bulk of our data (an intentional move as part of the larger multi-year research and design project). Data included 160 hours of observation documented with 183 pages of field notes and 191 photographs, six recorded interviews, two focus groups, notes from our design sessions, and other artifacts (e.g. librarian produced marketing, school district meeting minutes), collected over approximately 18 months.

By design, we emphasized an observational stance during the first phase of the project (roughly six months). That is, we limited any intentional introduction of new tools or resources and primarily conducted weekly observations and interviews with librarians and at our partner libraries to understand existing practice. This was in line with adoption of contextual inquiry (Beyer & Holtzblatt, 1998) as an overarching design
approach, where the central and primary commitment is understanding what individuals are doing within their existing spaces. In the second phase, we began to take our initial observations to develop prototype activities and resources that we then tested with librarians and youth patrons in focus groups and workshops, with the expectation that these would need substantial and continual revision. That phase has taken roughly a year of continuous work and just concluded prior to authoring this paper. A third phase is currently underway.

Data analysis
As a design-oriented paper, our goal is not to provide an extensive accounting of our analysis of all collected data. More traditional empirical studies are forthcoming. We do note, however, that our analysis activities have involved systematic coding of field notes and interviews using accepted methods associated with qualitative research (Saldaña, 2015) to inform our assertions. Given our goal of applying conjecture mapping and demonstrating how conjecture mapping can articulate changes in knowledge gained from design research, we present descriptive excerpts for illustration and justification rather than counted codes. Due to space limitations inherent in conference papers, we will only be reporting on some of our realizations and conjecture mapping.

Evolution of conjecture maps

The initial conjecture map
At the start of our design-based research project, we had a rudimentary mapping of how our design work would lead to a set of desired outcomes (see Figure 2). Namely, we collaborated with partnering library directors and school librarians who had expressed interest in having their libraries participate in the "Maker movement" that has been increasingly appearing in their professional literature (e.g., newsletters and journals) and communicated to them from various stakeholders or affiliates (e.g., district office coordinators for school libraries). The outcomes we hoped to realize were instantiations of high quality youth Maker programs created by our partner librarians and an increase in the amount of Maker programming. We knew that libraries, including those with whom we had partnered, already hosted youth programs and activities — whether it was gaming groups or afterschool homework clubs. This existing youth program development expertise was a key mediating process and a resource to enable the creation of Maker programs. As this was a new endeavor, we expected that our research team would have more direct involvement when we transitioned from observational to interventional work, with our university-based team members taking on more active development and facilitation roles for some Maker programs and helping to conceptualize program opportunities given our own extensive prior engagements with the larger Maker education research community.

Our expectation was that after some initial time investment in working with and observing the libraries and their personnel, we would be able to create a set of demonstration materials that embodied realistic enactments of Maker programs in small libraries. We also had the expectation that there were a number of existing digital informational and communications resources through which libraries learned about possible program ideas, such as social media and email lists within the local professional community that would enable knowledge sharing of different kinds of library-based programs. We hoped to discover and enact means for building upon those connective media. For instance, we assumed that social media use kept libraries abreast of what other libraries in the region were doing and could be reinforced to promote, model, and share ideas for Maker programs.

Given our expectation of existing use of informational and communications media, particularly within social media among librarians, and an ease with youth program development, our data collection approach was to initially observe existing library program development practice without any researcher intervention. That is,
we shadowed librarians at work, interviewed and observed them as they came up with youth programs, and attended and observed those youth programs and other routine activities at our partner libraries.

**Initial discoveries**
From our early observations and interviews, we were quickly humbled as project partners. The number of responsibilities and expectations placed on library professionals, whether they worked in a school or public setting, were well beyond what we had anticipated. In hindsight, this was naiveté on our parts by not having previously researched how librarians work and learn. For example, within the school library setting, there were constant demands on librarian time, with librarians taking on instructional planning, student advising, class supervision, assessment, financial management, school technology support, and fundraising above and beyond collections development and management. Observational interviews with school librarians often took place while things were momentarily quiet and the librarian was answering our questions while simultaneously shelving, assisting students, or preparing the library for a class of students arriving during the next period.

In the public library setting, there were demands on staffing, grant writing, customer service, report preparation, and community relations. In addition to these activities, this meant that youth programs were considered in fleeting moments when they were not already obligated to another service. Very limited staffing meant that there was one full-time employee in the school libraries and usually just one at the public libraries, given their small size. In public libraries, when there was the luxury of more than one full-time staff member, the individual who coordinated and facilitated youth programming had several other responsibilities beyond that. This is in contrast to larger libraries that have a dedicated youth services librarian.

**The need for “at-a-glance” program planning materials**
When we observed and inquired about youth program planning, we saw that the librarians who were most directly involved would conduct quick online searches for program examples, with social curating sites such as Pinterest mentioned regularly. Often stated as driving program constraints were what librarians had heard about or knew youth would be interested in doing and what would be relatively quick to prepare and familiar enough that they could execute the program given limited preparation time and budgetary resources. As we probed what led librarians to proceed with program design in this way, we found that it was largely driven by a desire to find new ideas quickly, to “see things in pictures,” and to adapt without extensive preparation on their part. For instance, one librarian relied on Pinterest for images of Harry Potter craft activities that could be adapted as a youth library program.

As we began some very early user tests of prototype demonstration materials with librarians in the form of print materials, images received the bulk of attention, large sections of text were ignored, and the librarians made active requests for more pictures of completed examples and how they worked. That led us to focus on highly pictorial materials and to explore strategies to represent completed examples and models in more immediate ways than what existing Maker education resources typically do. In this way, we discovered that the quick glance a librarian might take would reveal to them what could be possible and attainable.

**Variable images for desirable librarianship practice**
We also learned from our observational work that different librarians had very different images for what constituted a successful youth program. On the more conservative side, a public and a school librarian stated that libraries should be lively, noisy spaces during adolescent youth programs. We observed, and have noted elsewhere (Lee, Lewis, Searle, et al., 2017), that a librarian’s sense of librarianship practice – including what youth behaviors should be like, what purposes youth programs served, and what should be expected during a youth program – strongly influenced how different forms of youth program were designed and enacted by different librarians. Those who saw their work as maintaining and reproducing services in a self-contained way tended to have rigid programs that were not well attended by youth. Those who saw their work as reinventing the library and connecting to whatever resources were available in their immediate community had highly attended programs and were described very positively by vocal patrons and community members.

**Treatment of physical and occupied space in the library**
Finally, we observed that at programs with high levels of youth participation, the youth acted in ways that suggested a very different relationship with the library space and place than would be seen in a more typical
classroom. Libraries that hosted highly attended youth programs had youth who spoke out without raising their hands, would take off their shoes and walk barefoot in the library aisles, would sit with their feet on tables or sit on tables themselves, and play loud music and dance. One public library hosted teen programs after regular library operating hours. Adult patrons were not allowed, furniture was freely rearranged, and teens ran, laid on the ground, texted friends or played board and phone games, or found private nooks to sit and congregate with friends. This suggested to us that how the library space was understood—in terms of who should have access and what behavioral norms were acceptable—were important for youth to be motivated to actively attend and engage. This was confirmed in interviews with librarians. These observations led to a major revision of our conjecture map, described in the next section, as we more actively began to develop and enact new programs.

The second conjecture map
As we transitioned to design and implementation of Maker activities and away from strict, non-interventionist observations, one major change to our conjecture map was a change in its high-level conjecture (see Figure 3). Whereas before our conjecture had been about enabling Maker programs to take place in the library generally, we came to recognize that our partner librarians felt limited in their prior knowledge related to Maker programs and limited in their time to learn what they believed was necessary to support Making. This tied into a perceived idea of who gets to be a Maker and what activities constitute Making. In light of that, we internally and externally adjusted our expectations that the Maker programs offered in these libraries would appropriately be presented as low threshold, entry-level activities. We did not expect that the librarians would establish and lead robotics and coding clubs, and recognized that they often preferred to bring in someone else from their community (e.g., another teacher in the school for a school library or a representative from a community organization) for activities that went beyond what they felt they could learn and lead in a short amount of scattered preparation time. While supporting that, we also wanted to empower the librarians to lead programs.

Beyond changes to our high level conjecture, the embodiments of our design activities changed dramatically from what we had originally anticipated. In light of how we observed librarians curating program ideas and what kinds of materials they felt comfortable using, we began to develop visual guide program materials that were intended to encapsulate information about how a library Maker program could be sequenced, what materials were needed, and some fundamentals necessary for the creation of workable digital artifacts. For instance, a simple template for creating a basic paper circuit was necessary to establish some basic rules for connecting components into a functional circuit. When we tested templates that were available freely online, we discovered where novice users made errors (such as in tearing copper tape to make corners or creating short circuits near batteries) and made our own version of simplified materials that avoided some of these challenges.

We then found was that once we created these visual guide program materials, some librarians thought they should also be given to their youth patrons. This meant that some youth received materials explaining how library activities should be sequenced, how to address learning goals, how to communicate with youth patrons. In light of that, we began to deliberately separate materials for the librarian to plan the program, and materials for the librarian to copy and share with youth. This combination of librarian materials and patron materials was intended to provide new images for what could be done within the library, how the space within the library could be used, and decentralize expertise so librarians did not feel that they must present themselves as experts.

![Figure 3](image-url)  
**Figure 3.** The second conjecture map, with major changes highlighted in gray.

During our first enactments of newly developed Maker programs in the libraries, we observed immediately that youth had variable attendance. When a series of programs was offered over multiple days, some youth would show up for the first time on the third day and others would show up only on the first and last days. As such, we observed it would be difficult to design sequences of activities that became more advanced
over time because invariably, new youth who had never attended would arrive. We began to explore clearly identified ‘Newcomers stations’ so that first time and ‘drop-in’ program attendees could get started on their own. We also noticed that many youths who came complained that they did not know what to Make. After a few different attempts, and in consultation with participating youth, proposing a theme was recognized as an important program structuring element. For instance, if the theme was “Fun and Games”, attending youth enjoyed displaying how they could creatively embody that theme in their creations and what they could create that might push the boundaries of “games” (e.g., such as Pokemon Go vs. a board game). Also, we had observed some libraries had multi-week themes, such as “Marvel” films, that librarians reported resulted in increased youth attendance at programs. Building on that seemed to attract youth, according to librarians, and showed the range of ideas youth had for what could be made (e.g., a superhero mask vs. a weapon vs. a film scene).

Our intended outcomes from this conjecture map modification and for the subsequent six months of design and implementation work were: to emulate engaged youth participation comparable to what librarians reported as taking place during their most successful youth programs, to ease the librarian into being a comfortable facilitator and helper for students rather than feeling the need to be a content expert, and to use the successes of these experiences to empower librarians to create their own Maker-themed programs. We next describe two observations from enactments of library programs.

**Need for more initial investment from youth**

The new conjectures we established seemed more appropriate and tractable. Evidence for this included our librarians taking on more assistive (rather than leader) roles with youth patrons during program enactments, as well as strong program attendance. Still, some Maker activities and technologies had a much more enthusiastic reception from youth than others, including ones that the librarians and we had thought would be positively received. We thus began to see the need to increase youth involvement in initial program conceptualization. At some public libraries, teen advisory boards exist to get feedback on program ideas. Their members seem to also show heavy investment in library-based programs and are key figures in publicizing the programs and bringing new youth to the library. The literature on teen advisory boards is sparse, but working with them directly was fruitful for our design work and also empowered librarians to feel more confident in their decision making. We have since been working with our school librarians to encourage them to establish their own teen advisory boards, which is not a typical structure in school libraries. In addition, we introduced practices of intergenerational co-design (Guha et al., 2013) into researcher-librarian-youth program conceptualization.

**The ‘domino effect’ of participating in Making at the library**

As noted above, one librarian was more conservative and structured and tried to keep everyone at the same pace. In these instances, youth guide materials were supportive of the librarian’s preferred style of instruction. However, in more unstructured enactments, we saw librarians leave the youth guide materials out near the activity. Then, when these were used, one youth would figure out what to do, and then others would observe what that youth was doing and imitate him/her/them, and then occasionally expand on that in ways that would then be imitated by others in its new expanded form. In Maker learning activities, Blikstein (2013) noted a ‘keychain’ effect where youth would restrict themselves to making a single example when exposed to a new Maker technology, for example 3-D printing the same object (e.g., a keychain) repeatedly. In Blikstein’s work, the ‘keychain’ effect was viewed critically, rather than as an opportunity for learning. However, in our context, a ‘domino effect’ took place where guide materials we had designed were initially used by one youth, and observation of what that youth was doing and then personal modification contributed to other youth subsequently engaging in the Making activity. When new youth joined, they ignored the guide materials and opted to learn by watching. This social process of observation, imitation, and modification as a kind of vicarious learning was not one we had expected, and is now one we wish to account for in the design of library-based Maker programs. This process does not eliminate the need for youth guide materials, but has introduced different patterns and evolutions in youth participation.

**The current conjecture map**

We have begun a third phase of work involving another design iteration with new librarians and new cohorts of youth. This has also entailed revisions in our conjecture map and our related design work (see Figure 4). First, the split between program materials and youth guide materials seemed consistent with what librarians wanted and used. However, we observed that for some librarians, the youth guide materials took primacy and the programs became exercises in students working through those only. We also observed the aforementioned ‘domino effect’ where youth guide materials were used once by one or a few youth to get the activity started and then ignored by others.
Additionally, drawing from Learning Sciences innovations where science curriculum materials were enhanced to become educative curriculum materials (Davis & Krajcik, 2005), we began to modify the librarian program materials to become educative. We have systematically reviewed recently identified design principles for educative curriculum materials (Davis et al., 2017) and identified translations for the library setting. We are adopting those principles for the library context and are examining how well our embodiment of those supports changes how librarians view, enact, and facilitate educational programs.

Figure 4. The current conjecture map with changes from previous iteration highlighted in gray.

With respect to outcomes, the ‘domino effect’ described above alters our expectations for librarian facilitation. In part, we have come to see that a librarian’s sense of librarianship – including their view of how activities should be enacted in libraries and what role the librarian should play – still influences their facilitation. Changing the sense of librarianship is a long term endeavor, comparable to how teachers change their teaching practice. Still, we have observed modifications in the librarians, such as expanding the range of program offerings within library spaces and using teen advisory boards to help distribute expertise and support youth programs. Informally, we are seeing an increase in new library programs being initiated by our partner librarians using prior programs as inspiration. For example, we are seeing one partner school librarian encouraging the use of cardboard with students to engineer miniature skating parks and launching steampunk fashion making activities without our prompting and another has been making terrariums. As this next iteration of design and enactment of youth Maker programs in libraries unfolds, we will continue to examine if our conjectures are indeed plausible and attainable through our design work and revise and test more as necessary.

Conclusion

The goal of this paper has been to demonstrate and apply conjecture mapping grammar to iterative work with small town public and school libraries. We presented three iterations of our conjecture maps developed over 18 months of a design research partnership. In doing so, we note several things. First, while conjecture mapping pushes for specificity in hypothesized relations between the design, the learning setting, and the outcomes, the process of conjecture mapping also forces reflection on constraints. Many examples in Learning Sciences research that have reflected upon design have been in classroom settings. By moving to the library setting, a new set of constraints associated with the target population and demands on professionals working in this space have surfaced. The specific embodiments of design decisions were informed by these constraints, and the process of conjecture mapping prompted reflection on them. In conjecture mapping of new domains, we believe it is necessary and productive to identify and articulate constraints. That may not require additional components to a conjecture map diagram, but they are important parts of the design process.

Second, conjecture maps will undergo change over time, and design cases that make these changes visible will be instructive for the community. In our work, we have recognized where we were too general in our assumptions, such as what supports and exemplars would be adequate for use in library-based learning. Indeed, the need for more necessary supports in the process of iterative design has been documented elsewhere (Edelson et al., 1999). However, a push for final-form relations can mask the many decisions and influences involved in realizing an educational design. Like others (Svhila & Reeve, 2016), we believe that surfacing these helps the community appreciate and relate to the complexity of the work of doing design research.

Finally, conjecture maps push researchers toward articulating mediators within the system. For us, we have begun to articulate the kinds of entities and competencies that one ought to consider in our design space, such as how physical space is arranged or the sense of librarianship maintained by individual librarians. The precision required of conjecture maps encourages these articulations, and thus it seems to be a promising support for ontological innovations in design research (diSessa & Cobb, 2004).
References


Acknowledgments

This work was supported by Institute of Museum and Library Services grant number RE-31-16-0013-16.
Blocks or Text? How Programming Language Modality Makes a Difference in Assessing Underrepresented Populations

David Weintrop, University of Maryland, weintrop@umd.edu
Heather Killen, University of Maryland, hkillen@umd.edu
Baker Franke, Code.org, baker@code.org

Abstract: Broadening participation in computing is a major goal in contemporary computer science education. The emergence of visual, block-based programming environments such as Scratch and Alice have created a new pathway into computing, bringing creativity and playfulness into introductory computing contexts. Building on these successes, national curricular efforts in the United States are starting to incorporate block-based programming into instructional materials alongside, or in place of, conventional text-based programming. To understand if this decision is helping learners from historically underrepresented populations succeed in computing classes, this paper presents an analysis of over 5,000 students answering questions presented in both block-based and text-based modalities. A comparative analysis shows that while all students perform better when questions are presented in the block-based form, female students and students from historically underrepresented minorities saw the largest improvements. This finding suggests the choice of representation can positively affect groups historically marginalized in computing.

Introduction

In an effort to broaden participation in computing and give learners a more accurate sense of the field of computer science, and how it involves more than just programming, a new Advanced Placement (AP) course was created in the United States. The new course, titled AP Computer Science Principles (AP CSP), was taught nationally for the first time in the 2016-2017 school year and covers seven big ideas in computing: Creativity, Abstraction, Data, Algorithms, Programming, Internet, and Global Impacts. To help foster an inclusive learning experience, the course is programming language agnostic, allowing teachers to choose the technologies and programming environments for instruction. This means some learners go through the year-long high school course using block-based programming environments like Scratch, while students in other classrooms might use conventional text-based programming languages like Python. This presents a challenge for the organization tasked with creating a single written summative assessment to be administered nationally at the conclusion of the course: how do you create a written assessment for a computer science class when you do not know the programming language, or even the programming modality (text-based or block-based) that learners used? Further, because learners can use either modality, what can this assessment tell us about the role of programming modality towards the goal of broadening participation in computing? Given that a goal of the course is to broaden participation in the field, the assessment must reflect both the programming plurality welcomed in the design of the curriculum as well as the equity-oriented priorities of the course.

In response to this challenge, the Development Committee for the AP CSP course invented a custom pseudocode and included both text-based and block-based forms of questions on their exam (Figure 1). In this way, students are not rewarded or penalized for using one type of programming tool or another during the school year. However, there are concerns associated with this approach. By creating a new pseudo-language that no student has used during the course, it is unclear how students will perform. Further, many of the identified benefits associated with block-based tools are absent from the block-based form of the assessment. For example, the block-
based pseudocode does not use colors, present commands using natural language expressions, or differentiate block-shape by use. Further, the assessment only asks questions related to program comprehension, thus students are not asked to compose programs which is another a perceived affordance of block-based tools (Weintrop & Wilensky, 2015a).

The decision to both invent a pseudocode language that has two modalities and to present questions on the exam in both modalities is consequential for any student but especially given the goal of welcoming students from historically underrepresented populations to the field of computer science. If the questions asked on the summative exam do not align with the features of the learning environments that have been found to support students from historically underrepresented populations, then the course has the potential to perpetuate existing inequalities, rather than solve them. Understanding the outcomes from the assessment approach used for the written AP CSP exam, particularly with respect to the goal of broadening participation in the field of computer science, is the focus of this paper. Stated more explicitly, this paper answers the following research questions:

RQ 1: How do students perform on questions asked in an unimplemented text-based pseudocode compared to an isomorphic block-based pseudocode on a written computer science assessment?

RQ 2: How does modality (block-based versus text-based) affect students from historically underrepresented populations on a standardized computer science assessment?

To answer these questions, we designed a 20-question assessment that asks questions using both the block-based and text-based version of the pseudocode and embedded it as part of an AP CSP curriculum. Over 5,000 students took the assessment at schools across the United States. Using the responses, we are able to shed light on the stated research questions and advance our understanding of how the design of programming languages and assessments can impact learner outcomes and the goal of broadening participation in computing.

Literature review

Broadening participation in computing

Despite the growing presence of computing in society, the field of computer science still struggles with issues of underrepresentation of girls and African American and Hispanic learners (Ericson & Guzdial, 2014; Margolis, 2008; Margolis & Fisher, 2003; Zweben & Bizot, 2015). Efforts to increase enrollment among historically underrepresented minorities have taken a number of forms, including national curricular efforts, the creation of new programming environments that emphasize creativity and collaboration, and a wide array of out-of-school programs oriented toward engaging a diverse set of learners. For example, the Exploring Computer Science (ECS) curriculum was designed to broaden participation in computer science by emphasizing aspects of computing such as web design and data analysis and foregrounding the human and social aspects of the domain through culturally relevant curricula and equity-oriented projects (Ryoo et al., 2013). Another approach to introducing a broad range of learners to foundational computer science ideas is to integrate them into existing courses (Barr & Stephenson, 2011; Weintrop et al., 2016).

There are also a growing number of learning environments, technologies, and programs being designed to broaden participation in computing. Low-threshold programming environments have emerged, such as Scratch (Resnick et al., 2009) and Alice (Cooper, Dann, & Pausch, 2000), which use block-based programming interfaces to make it easier for novices to program with little or no prior experience. In these types of environments, programming is framed as a creative activity, allowing learners to create games and interactive stories that can easily be shared with others. Such efforts attract diverse learners that are historically underrepresented in computing (Kelleher, Pausch, & Kiesler, 2007; Maloney et al., 2008). Likewise, tangible computing has emerged as another pathway into computing that can reach audiences historically underrepresented in the field (Brady et al., 2016; Buechley & Hill, 2010).

Block-based programming

The last decade has seen a proliferation of applications utilizing blocks-based programming (Figure 2). Block-based programming leverages a programming-primitive-as-puzzle-piece metaphor that provides visual cues to the user about how and where commands can be used. Users compose programs in these environments by dragging blocks onto a canvas and snapping them together to form scripts. If two blocks cannot be joined to form a valid syntactic statement, the environment prevents them from snapping together, thus preventing syntax errors but retaining the practice of assembling programs instruction-by-instruction. Along with using block shape to denote use, there are other visual cues to help programmers, including color coding by conceptual use and nesting of blocks to denote scope (Maloney
Collectively, the features of the block-based programming modality and the environments in which they are situated contribute to learners perceiving block-based tools as easy to use (Weintrop & Wilensky, 2015a). A growing body of research is investigating how blocks-based programming shapes learners’ conceptual understanding of computer science concepts and emerging programming practices. For example, researchers have documented a number of habits learners develop while working in blocks-based tools, such as an emphasis on bottom-up programming where learners focus on using specific blocks (Meerbaum-Salant, Armoni, & Ben-Ari, 2011). Likewise, research is documenting how novices learn with blocks-based tools; identifying developmentally-appropriate content and misconceptions learners may develop in blocks-based languages (Franklin et al., 2017; Grover & Basu, 2017; Seiter & Foreman, 2013). Finally, emerging research on comparative studies between block-based and text-based tools shows block-based programming enables students to complete assignments faster (Price & Barnes, 2015) and score higher on content assessments (Weintrop & Wilensky, 2017).

Representation and learning

Research in the Learning Sciences is revealing the ways that the representational infrastructure used in a domain can impact learning and conceptualization of that domain. diSessa (2000) calls this material intelligence, arguing for close ties between the internal cognitive process and the external representations that support them: “we don’t always have ideas and then express them in the medium. We have ideas with the medium” (diSessa, 2000, p. 116, emphasis in the original). These symbolic systems provide a representational infrastructure upon which knowledge is built and communicated (Kaput, Noss, & Hoyles, 2002). For example, Sherin (2001) investigated the use of conventional algebraic representations as compared to programmatic representations in physics courses and found that different representational forms have different affordances with respect to students learning physics concepts.

Wilensky and Papert (2010) give the name structuration to describe this relationship between the representational infrastructure used within the domain and the understanding that infrastructure enables and promotes. While often assumed to be static, Wilensky and Papert show that the structurations that underpin a discipline can, and sometimes should, change as new technologies and ideas emerge. In their formulation of the idea of structurations, Wilensky and Papert document a number of restructurations, shifts from one representational infrastructure to another, including the move from Roman numerals to Hindu-Arabic numerals (Swetz, 1989), the use of the Logo programming language to serve as the representational system to explore geometry (Abelson & diSessa, 1986), and the use of agent-based modeling to represent various biological, physical, and social systems (Wilensky & Rand, 2014). This work highlights the importance of studying representational systems, as restructurations can profoundly change the learnability, power, and communicability of ideas within a domain.

Methods

Context: The AP Computer Science Principles course and exam

This study took place in classrooms during the inaugural year of the AP CSP course. The year-long course focuses on the big ideas of computing and culminates with a 2-hour, 74-question multiple choice exam. Roughly 20% of a student’s overall score on the AP Exam is based on their responses to programming questions. The programming questions on the written exam include a mix of block-based and text-based pseudocode questions of the form presented in Figure 1. The assessment used in this study was administered as part of Code.org’s CSP curriculum (http://code.org/csp). Code.org’s CSP curriculum is a full-year course that introduces high school students to the foundations of modern computing and prepares them for the AP CSP Exam. The course employs curricular and pedagogical strategies that promote equitable teaching practices to support both new-to-computer-science students and teachers. Two of the five units in the curriculum involve programming in the JavaScript language through
Code.org’s App Lab environment (Figure 2d, http://code.org/applab), which allows students to construct programs in both block-based and text-based modalities and includes various affordances to help novice programmers.

The AP pseudocode and our assessment
As shown in Figure 1, the pseudocode created for the AP CSP exam draws inspiration from professional programming languages but includes features and keywords that make it distinct from any existing language. The pseudocode includes keywords for the following topics: Assignment, Display, and Input; Arithmetic Operators and Numeric Procedures; Relational and Boolean Operators; Selection; Iteration; List Operations; Procedures; and Robot commands. The block-based and text-based versions of the pseudocode are the same with a number of exceptions, notably, in the block-based form expressions are wrapped in a rounded-edge rectangle and scope is denoted with nested rectangles, replacing the `{}`s used in the text form. The language consists of 23 keywords including looping constructs (e.g. FOR, REPEAT), conditional operators (e.g. IF, ELSE), and list functions (e.g. REMOVE, LENGTH).

The assessment used in this study asked questions using the block-based and text-based form of the AP CSP pseudocode and followed the design of the Commutative Assessment (Weintrop & Wilensky, 2015b). Each question on the assessment began with a short program presented in either the text-version or block version of the AP CSP pseudocode (e.g. Figure 1) and then followed by the question: “What will the output of the program be?” The assessment was comprised of 20 multiple-choice questions covering five programming topics: variables, loops, conditionals, functions, and program comprehension. For each of the five topics, the assessment asked two questions in the block-based pseudocode form and two in the text-based pseudocode. This counter-balance design ensures that every student answered at least one question for each concept in both forms of the pseudocode.

Data collection and participants
The AP pseudocode assessment was included at the end of the App Development module of the Code.org AP CSP curriculum under the label: AP Pseudocode Practice Questions. The assessment was a supplement to the official course materials. Teachers were notified of its existence via a monthly newsletter that was sent to Code.org users that identified as having an active section of the course. The Code.org curriculum is run through a content management system that tracks individual student as well as classroom progress through the curriculum. This platform was used to administer the assessment and to collect and store student responses. At the beginning of the course, students are asked to create a profile, which includes optionally self-reporting their gender, age, and race. The responses collected by Code.org were de-identified and then shared with researchers. For each student, the data included a full set of responses to the survey questions along with birth year, gender, and race.

The dataset for this project consists of 5,427 students and over 105,000 individual questions responses. The sample was comprised of 1,218 (22.4%) female students, 3,198 (58.9%) male students, and 1,011 students (18.6%) who chose to not provide gender information. Of the 5,427 students, 1,040 (19.2%) learners self-identified as being from an underrepresented minority in computing (URM), while 2,199 (40.5%) of students were classified as not URM and 2,188 (40.3%) of student did not self-report their race. For this work, students that self-identified as Black, Hispanic, LatinX, Native American, or Pacific Islander were categorized as being from a URM. Finally, a majority of participants were between the ages of 15 and 18, which corresponds to the four years of American high-school (9.6% 15 years old; 22.5% 16 years old; 30.7% 17 years old; 27.2 % 18 years old).

Findings
This section presents our analysis of the responses to the AP CSP assessment beginning with overall results then looking at how modality affected learners within racial and gender groups and how outcomes differed by modality across groups. For each calculation, we only include students who answered all of the questions for the calculation being run and provided the necessary demographic data.

Overall findings
Of the 5,427 students who answered at least one question, 4,762 of them that provided responses for all 20 questions on the assessment with a mean score of 16.15 points out of a possible 20 (SD 3.9). Of these 4,762 students, 758 students (15.9%) scored a perfect 20 out of 20, while 369 students (7.7%) correctly answered less than half of the questions. The mean correct response rate for girls was 15.79 (SD 3.776) while boys average score was 16.10 (SD 3.99). An independent samples t-test reveals these two scores to be statistically different from each other $t(3933) = 2.20, p = .03, d = .08$. Using Cohen’s (1992) guidelines, the effect size $d$ (calculated using the pooled standard deviation) is interpreted as a small effect. This result shows boys outperforming girls on the
assessment and that the difference in scores between the two groups was statistically significant but relatively small.

Shifting our analysis to students who self-reported as being from an underrepresented minority, we find a similar pattern, but a larger effect size. The mean correct response rate for URM students was 15.26 (SD 4.17) while those reporting as not URM scored an average of 16.64 out of 20 (SD 3.49). An independent samples t-test shows these two groups to be statistically significantly different, \( t(2878) = 9.2, p < .001, d = .36 \). This is interpreted as a medium effect. This finding shows URM students scoring significantly lower than non-URM peers, which, along with the previous finding of boys outperforming girls, matches prior work related to achievement and race and gender.

![Figure 3](Image)

**Figure 3.** Average scores on the assessment presented as comparisons both within and between population groups.

**Differences by modality, within population group**

The first research question we pursue in this work investigates how modality affects learners’ ability to correctly answer questions on a written computer science assessment. The assessment we designed included 10 block-based questions and 10 text-based questions, providing a dataset that allows us to compare performance between modalities within subjects and within groups. Overall, students scored an average of 8.5 out of 10 on block-based questions (SD 1.9) and 7.7 out of 10 on text-based questions (SD 2.2), a difference that is statistically significant \( t(4761) = 38.14, p < .001, d = .40 \) (left-most column of Figure 3). This means students performed better on the block-based question than the text-based questions, a finding that matches prior research looking at students ability to answer written questions by modality (Weintrop & Wilensky, 2015b). However, this result is slightly surprising as the block-based format used in this assessment (Figure 1) lacks many of the features the literature has identified as making block-based tools easier to comprehend, a fact that we will return to later in the discussion.

We now break the population down to see if this trend is consistent across subgroups. Looking at gender, we find that both male and female students perform significantly better on block-based questions: Female \( t(1081) = 20.25, p < .001, d = 0.62 \) and Male \( t(2852) = 29.34, p < .001, d = 0.55 \) (columns 2 and 3 in Figure 3 respectively). In the case of female students, this resulted in a .92 point improvement on average while males experienced a .85 increase in mean score. For both genders, there was a medium effect size for performance by modality. Shifting analytic focus to race, we see a similar pattern with both URMs and non-URM students performing significantly better on block-based questions versus text-based questions: URMs \( t(878) = 18.20, p < .001, d = 0.61 \) and non-URMs \( t(2000) = 23.74, p < .001, d = 0.53 \) (columns 4 and 5 in Figure 3 respectively). For participants that self-identified as a URM, there was, on average a full point difference between mean block-based score and mean text-based score. Non-URM students saw an average score improvement of .76 points between block-based and text-based questions. Like with gender, both statistical differences are found to have a medium effect size. Taken together, this analysis shows all students perform better on block-based questions, however, looking at the difference in effect size, students from historically underrepresented populations see a greater benefit when questions are asked in the block-based form. This suggests that using block-based programming is a useful approach to supporting learners from underrepresented groups with the goal of broadening participation in computing. Having looked at differences within populations, we now shift our focus to a comparison across populations.

**Differences within modality, across population groups**
The previous analysis established that there are differences in performance patterns between block-based and text-based questions. In this section, we conduct a similar analysis, this time looking comparatively across groups to try and identify the sources of differences identified above. We begin by looking at differences by gender. Comparing student performance on block-based questions by gender using an independent sample t-test shows a not statistically significant difference between genders: $t(3969) = 1.6, p = .12$ (column 6 in Figure 3). If we run the same comparison looking only at text-based questions, we do find a statistically significant difference between gender: $t(3981) = 2.2, p = .03, d = .08$ (column 7 in Figure 3). Taken together, these findings show that the source of difference in performance by gender reported above is due to results from the text-based questions. In other words, there is no difference in score by gender on the block-based questions, but a significant difference in scores on the text-based questions. This suggests that one potential way to shrink the gender gap would be to only use the block-based modality.

If we look at comparative differences by modality between URM and non-URM students, we find a different pattern. The average score for URM students on text-based questions was 7.1 out of 10 (SD 1.4) compared to 7.9 (SD 2.0) for non-URM students with a similar difference in average scores being found for block-based questions: URM students averaged 8.1 out of 10 (SD 2.2) and non-URM students scored an average of 8.7 out of 10 (SD 1.8) (the two right-most columns in Figure 3). For both block-based and text-based questions, these differences are statistically significant: block-based questions $t(2909) = 7.95, p < .001, d = .30$ and text-based questions $t(2917) = 9.88, p < .001, d = .38$. The difference in patterns with these results compared to the gender differences observed in the previous paragraph suggests we would not expect the same erosion of performance difference between racial groups if assessments shifted to use the block-based modality exclusively, an idea we further explore below.

Discussion

Potential explanations for these differences

One of the surprising findings from this research is the presence of significant differences between block-based and text-based questions despite the block-based pseudocode not having many of the features that learners have identified as being most helpful. Prior research shows that students identify things such as ease of composition, meaningful shapes and colors, natural language expressions, and ease of browsability as key features of block-based programming tools that helped them learn to program (Weintrop & Wilensky, 2015a). However, none of those features are present in the block-based pseudocode used in this assessment. This suggests there are other factors at play that can explain the observed differences. One potential option is that the block-based visual cues that are present (like the nesting of scope) does play a role in helping learners comprehend the program. This supports prior research that has found navigating scope within a program to be one strength of block-based programming environments (Weintrop & Holbert, 2017), as well as work showing students struggle with unfamiliar syntax (Stefik & Siebert, 2013).

Another possible explanation has less to do with the specifics of the text-based syntax or the rendering of the block-based interface, but instead, has more to do with the affective and perceptual aspects of the questions being asked. Students who are intimidated by text-based programming, or learners who have spent more time working in block-based tools, may feel more comfortable trying to decipher programs written in the block-based modality. In this way, it is less the specifics of the representation that are contributing to these results, but instead, the larger cultural or social dimensions of the representations. Unfortunately, this study does not have the data sources necessary to verify this explanation, as will be discussed later in the section on limitations and future work.

A third potential explanation links the results on the two forms of the pseudocode questions with the programming modality used in class. It is conceivable that classes explicitly designed to broaden participation and have greater numbers of female and/or URM students were more likely to have used block-based tools. At the same time, schools in which male and non-URM students have historically enrolled in computer science may have used this new course as preparation for later computer science instruction and relied on text-based tools. In other words, the results presented above are due to differences in what populations were using what type of programming environment. While we do not have the data from this study to refute this explanation, prior work has not found strong coupling between the modality used for instruction and student performance (Weintrop & Wilensky 2015a, 2017).

Implications of these findings

There are a number of implications that flow directly from the findings presented above. Taken collectively, the presented findings suggest that block-based programming can play a productive role in supporting learners from
historically underrepresented populations. This conclusion stems from the finding that female and URM students particularly benefit from the use of block-based questions. Thus, we would expect that if the exam were administered using only block-based questions, the gender gap would shrink, resulting in female students ending the class with a higher grade and potentially increasing the likelihood that they sign up for future computer science courses. There is still much work to be done before we can prove that the transition to block-based assessments produces these outcomes, but the data from this study suggest it is a promising avenue for broadening participation.

A second potential implication of these findings relates to the endpoint of computer science instruction. Historically, computer science education instruction in the United States has had the built-in assumption that text-based programming using a professional language (e.g. Java) was an essential component of early computer science instruction. With the emergence of block-based tools, and other non-professional programming contexts in which authentic computer science can take place, this text-based programming endpoint of computer science instruction is beginning to be challenged. Can someone become proficient in computer science without learning text-based programming? With the introduction of the AP CSP course in the United States, the answer is starting to be yes, but there is still work to be done to completely realize this new endpoint of computing education. The decision to teach an AP course and assess student understanding of computer science entirely in block-based language is a towards new computer science endpoints distinct from text-based programming in professional languages.

Limitations and future work
It is important to note the limitations of this work. For example, while we know some basic information about the participants, like gender and age, there is much we don’t know, such as prior programming experience, which would impact the results presented above. Likewise, we know very little about students’ classroom experience (e.g. the curriculum followed, the programming tool(s) used, and teacher’s prior computing education experience). All of these factors influence the data presented in this study and thus constrain the claims that can be made. A second limitation is that we only have a single, quantitative data source. Ideally, we would have also observed these classrooms to better understand the nature of the instruction as well as interviewed students after taking the survey to gain further insight into how students are making sense of the role the representations in the assessment. We are in the process of designing a second iteration of this study where additional contextual data will be collected to address this limitation.

Conclusion
The push to broaden participation in computing has resulted in numerous computing education initiatives, often championed and implemented with little or no prior research being conducted on the key elements of the program. In the case of the new AP CSP course, the decision to support a plurality of programming tools was informed by research, but the novelty in how the administering body chose to assess student learning led to uncertainty with respect to whether or not the final exam would help address issues of equity and underrepresentation. The analysis presented in this work found that the use of block-based pseudocode programming questions supported learners from historically underrepresented populations, suggesting that this evaluation choice may help the course achieve its stated goals of broadening participation in computing. This analysis advances our understanding of the role that representation can play in supporting diverse students in succeeding in computer science and contributes to the growing knowledge-based of pedagogical strategies and design techniques that can help all students succeed in computing.

References


Towards a Cognitive Ecological Framework in CSCL

Marcela Borge, The Pennsylvania State University, mbs15@psu.edu
Emma Mercier, University of Illinois at Urbana Champaign, mercier@illinois.edu

Abstract: The field of computer-supported collaborative learning uses a range of theories, grounded in different epistemic frameworks, to conduct research and design learning opportunities. While some studies use multiple theories, and attempt to bridge the frameworks, most use a single theory, considering learning from just one perspective. We argue that this can lead to a reduction in what we can understand about learning from each study, and has implications for the quality of designs and limits the possibility of uptake of CSCL tools beyond the specific research context. Drawing on other fields, we identify the potential and need for an ecological framework to inform research and design work, and argue for its importance in the development of the CSCL field (1).

Introduction

“Another success of the past two decades of theory-based research can be seen in the evolution of theories and models themselves—a move away from a major focus only on individual behavior change and toward broader multi-level behavior and social change models. By the late 1980s, the limited reach and staying power of even our most effective individual health behavior interventions, based on theories emphasizing intrapersonal and interpersonal determinants of health behaviors, made it clear that an exclusive reliance on individually oriented interventions would be inadequate to achieve our pressing population health and health care goals. These failures led to a fundamental “paradigm shift” in our understanding of what the targets of effective interventions needed to be, not just individuals but the broader contexts in which they live and work. This shift fueled the rise of ecological models of health promotion that have guided the development of powerful interventions in public health and health care arenas.” - Orleans, 2008

Many fields have made paradigmatic shifts in how they design and evaluate interventions, which have resulted in changes in the theoretical frameworks they use to guide and interpret research. In fields that emphasize the design of tools and interventions, theoretical frameworks have evolved from linear to more systemic conceptual models to guide their understanding of complex social phenomenon (Carroll, 2009; Engeström, 2000; Sallis, Owen, & Fischer, 2015). This shift is particularly necessary for unpacking problems that include social factors. Unpacking problems associated with decreasing bee populations is one example. Bees are social animals, so to understand why their populations are dwindling, it is necessary to identify possible environmental causes of bee death, the ways that individual bees deal with such environmental stressors, and how these changes affect the social dynamics and long-term outcomes of a colony (Perry, Søvik, Myerscough, & Barron, 2015). In other words, it is an ecological problem that has to be examined within an ecological framework.

In Computer Supported Collaborative Learning (CSCL), we aim to understand learning as it occurs in a socio-technical context. There are different theoretical frameworks that guide our research. Influenced by different historical perspectives on cognition and learning, each framework focuses on a small aspect of the learning system and have different views on where cognition occurs and how learning happens (Greeno, Collins, & Resnick, 1996). Differences between these perspectives have fueled debates between researchers on where cognition occurs and what counts as learning. However, there is ample evidence to support the claim that cognition is a multilevel, systemic phenomenon. As such, selecting only one framework to guide research limits our ability to understand learning on a broader scale and introduces bias into our research, as we commit sampling errors and ignore other possible explanations for research outcomes. Given the complex nature of computer-supported collaborative learning, we argue that it is necessary for CSCL to begin moving towards a cognitive ecological framework in order to better inform how we examine learning, trade-offs associated with different technologies and interventions, and our evolving understanding of how learning happens in social contexts.

Different theoretical frameworks to examine learning processes

The Learning Sciences, and CSCL as a part of the Learning Sciences, emerged from multiple disciplines and epistemic traditions, integrating a range of theoretical frameworks from which to study learning (see Table 1).
<table>
<thead>
<tr>
<th>Theory</th>
<th>Theoretical Assumptions</th>
<th>Unit of Analysis</th>
<th>Most Common Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Processing</td>
<td>Learning occurs as a function of psychological and biological mechanisms within the brain.</td>
<td>Individual processes occurring within the brain.</td>
<td>Individual measurement of acquired knowledge and skills: multiple choice tests, essay, and short answer; standardized tests.</td>
</tr>
<tr>
<td>Constructivism</td>
<td>Learning occurs as a function of individual construction of knowledge based on previous experiences and existing knowledge</td>
<td>Individual processes occurring between individuals or between individuals and objects.</td>
<td>Individual measurement of cognitive growth: assessments of conceptual change, performance tests in authentic contexts.</td>
</tr>
<tr>
<td>Social Constructivism/Sociocultural</td>
<td>Learning occurs as function of the gradual internalization and appropriation of cultural norms, expectations, practices, and value systems.</td>
<td>Collective processes occurring between individuals, between individuals and objects, within groups, or within communities.</td>
<td>Individual measurement of changes in discourse patterns, identity, social practices or artifact use.</td>
</tr>
<tr>
<td>Group Cognition</td>
<td>Learning occurs at the level of the group as individuals externalize individual thought through language and create new shared understanding; group cognition</td>
<td>Group processes occurring through language or through the creation of shared knowledge artifacts.</td>
<td>Collective changes in discourse patterns, social practices, or artifact use.</td>
</tr>
</tbody>
</table>

These frameworks range from those that take the individual learner as the unit of analysis to those that focus on the group. The richness that these different perspectives bring allow us to develop a nuanced and complex understanding of learning at multiple levels of analysis. However, concerns have been raised about the constraints of these theoretical stances, and how research and development of innovative learning environments may be stalled or stymied by these constraints (Akkerman et al., 2007; Jarvela, et al., 2010). The epistemological foundations of different theories are at odds, and therefore, for researchers who are attentive to such features, place roadblocks when trying to address learning at multiple levels in a single design research project or study.

Webb (2013) describes an information processing approach to understanding the outcomes of collaborative learning. This approach focuses on individual learning, suggesting that students learn through actively processing information while engaging in collaborative activities. From this perspective, individual cognition is what leads to the formulation and presentation of ideas; individual cognition is necessary for explaining, actively listening, and responding to a collaborator and can be positively impacted by collaborative learning. This approach, which focuses on knowledge in the individual’s mind, closely relates to constructivist theories of learning (e.g. Piaget, 1965). These theories pose that learning happens as a result of experiencing disequilibrium, the constructivist, cognitive processes of assimilation and accommodation, which may be caused by peer interaction. Studies that take this approach tend to focus on the individual learner, comparing pre- and post-test scores given to each participant (e.g. Meier, Spada, & Rummel, 2007; Webb, Nemier, & Zuniga, 2002).

A socio-constructivist theory of collaborative learning, places increased emphasis on the role of interaction with more-expert peers and the societal use of tools and language to pass down knowledge through generations (Golbeck & El-Moslimany, 2013). While Vygotsky maintained a focus on the individual learner, the emphasis differs from Piaget in that Vygotsky takes into account the need for a learner to “master the items of cultural experience” and “the habits and forms of cultural behavior” (Vygotsky, 1929, p 415). Interpretations of Vygotsky’s theory range from those that emphasize the role of the expert in learning and the zone of proximal development (e.g. Wood, Bruner, & Ross, 1976) and others that emphasize that learning is an act of increasing participation in the cultural norms and tool use of a community (e.g. Hakkarainen, et al., 2013).
remain close to a Vygotskian perspective often examine how learners at different stages of development or experience perform during interactions (e.g. Azmitia, 1988; Schmitz & Winskel, 2008). Studies that take a broader socio-cultural approach also examine how students co-construct knowledge, regardless of their prior experience or development level, and consider the use of tools in knowledge building. These studies may use individual learner outcome measures, but often make attempts to consider the non-independence of the data; these studies look at the quality of interaction, often attempting to tie it to learner outcomes (e.g. Barron, 2003; Mercier, 2017).

Group cognition theory alters the focus from the individual learner to the group interaction. Described by Stahl (2013) as a post-cognitive approach, where the nature of interaction among group members produces knowledge that cannot be attributed to any individual contribution nor understood as a sequence of contributions from the individuals in the group. In this way, group cognition shares similarities with psychological theories of macrocognition, which is defined as "the process of transforming internalized knowledge into externalized team knowledge through individual and team knowledge-building processes (Fiore, et al., 2010). These theories of group cognition are rooted in theoretical perspectives that perceive cognition as contextually situated and distributed among individuals, tools, and artifacts that are part of the sense-making processes (Fiore & Schooeler, 2004; Stahl, 2006). Thus, from this perspective, collaborative learning can be seen as an emergent property of the group, not as something that can be attributed to an individual. Assessment of learning from a group cognition perspective is often ethnographic in nature, focusing on analyzing the processes of learning that occurred during the interaction, rather than relying on post-test learning measures of individuals (e.g. Mercier, et al., 2013).

How different frameworks create tensions in the field
These different frameworks work well as a means to deeply understand learning at specific levels of interest. Information processing and constructivist theories have greatly added to our understanding of individual sense-making and knowledge building. Socio-cultural and socio-constructivist theories have helped us to recognize the importance of dialogic forms of learning, how cultural values and expectations impact sense-making activity and learning outcomes, and the importance of situated practice. Group cognition has helped us to develop a better understanding of how groups learn as it pushed us to examine verbal interaction as a form of shared thought and show how learning and cognition can occur at the level of the group. Each theoretical perspective has implications for how we measure learning outcomes and design learning contexts for individuals, groups, or communities of learners (Greeno, Collins & Resnick, 1996). However, prioritizing one of these frameworks over another and focusing solely on a singular level of learning activity is problematic.

While using research designs grounded in a single theoretical perspective is relatively unproblematic, difficulties arise when we begin to ask questions such as why collaborations are successful or consider designing CSCL activities that we wish to implement in classroom contexts. Multiple studies (e.g. Barron, 2000; Kapur & Kinzer, 2007; Webb & Mastergeorge, 2003) report that their findings explains some of the variance that emerges between groups, but few studies account for all of the variance. This limitation results in a patchwork of findings that provide some insight into the nature of successful collaboration, but does not give a full picture of what is occurring at the multiple levels of learning and interaction that occur during collaborations.

One particular example is the classic study by Roschelle (1992) in which he explored the central role of knowledge convergence to collaboration. He argued that the co-construction of knowledge is a process through which individuals converge on a shared meaning, making sense together, and as such, co-create knowledge at the group level, while developing an individual understanding of a phenomena. He found that neither the Piagetian concept of the individual’s cognitive development, nor the Vygotskian emphasis on the need for a more knowledgeable peer can account for the type of learning that went on between students during the collaboration that he observed, and that a more nuanced understanding of learning and knowledge across levels is necessary to account for this type of collaborative learning. He argues that there is little doubt that each student was engaged in individual sense-making or cognitive processes, but that there was also evidence for joint cognition, and that some of the learning occurs during the interactive processes of the dyad. The group was neither merely providing incentive for accommodation and assimilation at the individual level, as Piaget would suggest, nor was there a more knowledgeable peer supporting the learning, as both students were making sense of the information for the first time. Thus, he argues for a more complex theory of collaboration to account for the observed processes.

Group cognition as a theoretical framework, addresses some of Roschelle’s concerns, but does not allow for the central nature of individual sense-making during the co-construction of knowledge that Roschelle reported. Group cognition alleviates the reliance on a more knowledgeable peer, as is central to Vygotskian theory, and provides a framework to account for the knowledge creation that happens between people, when all are novices or when they encounter a new problem. However, there is no place in this theory for individual cognition, and no way to consider what the individual may learn from participating in such an activity.
Salomon & Perkins (1998) argue for the importance of recognizing the synergistic interaction between the individual and social aspects of learning. Reviewing the literature that situates the individual mind and social interaction within different conceptions of where information processing lies, they suggest that it is necessary from a design standpoint to understand how these two processes may interact and, rather than foreground one or the other, account for how they play equal parts in learning as we design instruction. Similarly, Jarvela, Volet, & Jarvenoja, (2010) also note a need to move beyond the situative and cognitive divide when we consider the role of motivation on collaboration. They argue that social and individual processes occur simultaneously and in interaction with each other, and that by limiting our focus to one level of analysis, we cannot fully understand the nature of the phenomena we are studying.

Akerman and colleagues (2007), argue for a need to bridge the divide between sociocultural and cognitive approaches to collaboration, but note that few studies successfully manage this. The studies that successfully cross the boundaries between both individual cognition and group cognition in their analysis, point towards the need for the field to develop theories that allow for such analyses without the epistemic divide central to these two approaches. However, basic differences in what counts as knowledge, hinder our ability to do this successfully. This suggests the need for a more coherent framework – not just merely the juxtaposition of multiple frameworks in our analysis, but a new framework that allows for a broader conception of knowledge and collaborative learning.

In support of a multi-level, systemic model of cognition and learning

Regardless of our particular theoretical stances most can agree that there are many different types of knowledge that can be tacit or explicitly held. There are also many types of learning: cognitive, metacognitive, sociocognitive, and socio-metacognitive learning. For example, let us examine what students could learn when they play an online, collaborative science game. Students could develop their knowledge about the features and functions of the game, they could develop knowledge about the scientific content contained within the game, or they could develop knowledge about different strategies that can be used to overcome obstacles in the game. However, students could also be learning other things as well. Metacognitively, they could be learning how to monitor and regulate their playing processes to achieve desired gaming outcomes. Sociocognitively, they could learn how to problem solve with other students in order to figure out how to overcome difficult aspects of the game. They could also learn that their peers react negatively to teasing during gameplay. From a meta-sociocognitive perspective, they could also learn about desired social and emotional practices and how to monitor and regulate their own and others’ social interactions to promote a more positive learning environment.

There are also many variables that can influence all these forms of learning and interfere with knowledge development. These variables include identity formation, beliefs, values, interests, proclivities, skills, and the ongoing outcomes of interactions between our unit of analysis (the individual, the group, or the community) and other environmental factors such as access to food, educational resources, mentors, social acceptance, access to ongoing positive educational experiences, stable homelife, fear of violence, etc.

It is also possible that one form of learning could interfere with another. A computer game may help students develop specific in-game skills and knowledge, but over long periods of use could interfere with social forms of learning as the individual becomes immersed in gameplay thereby reducing social discourse and interactions with others. Sherry Turkle’s work points to growing evidence that supports this relationship. For example, Turkle (2016) discusses how social media and cell phones may increase our factual knowledge about where people are, what is going on in the world, and what people like, but diminish our ability to develop the social skills to connect with people who are different than us or engage in meaningful conversations.

Another possibility is that learning at one level of analysis (the individual, the small group, or the community) can help or hinder learning at another level. For example, if an individual develops knowledge that others lack, a small group may ignore that person’s contributions and fail to build on this unique knowledge. This is because groups have tendencies to build on shared knowledge over unique knowledge (Stasser & Titus, 2003). Individuals who have not learned how to share authority may also dominate group conversation and interfere with the entire group’s learning and performance (Barron, 2003; Borge & Carroll, 2014; Woolley et al., 2010). Groups can also develop knowledge in a situated context, but fail to understand the series of individual contributions and decision-making processes that led to that knowledge development. As a result, individuals would not learn the whole of the knowledge and the community may be unable to reproduce it. Organizational learning, learning at the level of the larger community, is also subject to many problems. Individual and small group level characteristics can prevent learning from occurring across the community. For example, toxic individuals or small groups that make others feel devalued can create an environment that feels psychologically unsafe. Such environments prevent individuals or small groups from or sharing errors, admitting when they do not understand,
or contributing viewpoints, thus interfering with organizational learning that occurs across individuals and small
groups in the community (Edmondson, 1999).

What these collective studies imply is that learning and behavior is not linear, or even nested. It is
ecological, meaning it is a multilevel, multisystemic model. Learning occurs as part of interrelated, social systems
(Cole & Engestrom, 1993; Jonassen & Rohrer-Murphey, 1999). Learning occurs in different ways at different
levels and can be affected by different factors within levels and interactions between factors across levels. If we
accept that learning is an ecological phenomenon, one that is dependent on a relationship between different people,
groups, communities, and their physical surrounding, then we can come to understand that an instance of learning
within such complex systems cannot be fully understood by examining that instance in isolation from its
interconnected parts. Many other fields have come to recognize this problem and have moved to more ecological
models to guide their developing understanding. For the field of CSCL to continue moving forwards, it is
necessary that we consider moving to an ecological model of learning.

The effects of ignoring multi-level, systemic models
Many fields have recognized that ignoring the multilevel systemic nature of human activity is problematic for the
design of tools to help people improve their activities. The fields of Computer Science and Public Health are two
eamples. Both of these fields began making sense of human learning and behavior from an individualistic
information-processing perspective. They designed tools and interventions that were uninformed by the activity
systems that their subjects belonged to, but came to recognize that this approach was largely ineffectual and costly.

When computer science was emerging in the 1970s most computational technologies were developed by
computer scientists for computer scientists and enthusiasts. When personal computing became more prominent in
the 1980’s the need to create technologies for computer novices became necessary. As a means to address this
need, many universities and research institutions began examining how everyday people completed tasks, how
computers could help to simplify these tasks, and how computers could be built to be fairly usable by these
workers. This is how Human-Computer Interaction (HCI) emerged as a field: strongly tied to cognitive science
and human factors research and mainly confined to the study of individual computing activity (Grudin, 2008).

HCI changed over time as human activity and technology creation co-evolved. As technologies became
more prevalent and intertwined within different aspects of human activity, the ways that people worked, played,
and communicated also changed. Overtime, the field of HCI began to recognize the systemic and ecological nature
of human behavior, learning, and artifact use. As a result, the theories that HCI used to inform design gradually
shifted from theories that explain the individual to theories that explain activity systems and the interplay between
learning, behavior, and emotion (Carroll, 2009). This shift has helped HCI to grow and for various fields outside
of Computer science to recognize the importance of Human-Centered Design.

Similar methodological progressions can be found in health medicine. In their book on health and human
behavior, Glanz, Rimer, and Viswanath (2008) capture the increasing complexity of health intervention research.
The history of the field shows striking similarities to the growing complexity and theory application described in
HCI. Unlike HCI, the emphasis is not on the development of products, but the creation of interventions to change
existing health practices. In the 1980s, most educational health interventions were informed by individual theories
of learning and behavior, but no matter how theoretically informed these interventions were, they remained largely
ineffective. These failures led the field to conceptualize human behavior and interventions as existing within an
ecological system. Glanz et al. (2008) discuss how this shift resulted in the design of educational interventions
that targeted and evaluated change at multiple levels of analysis: the individual level, interpersonal level, and the
community and organizational level. This approach, they say, helped the field to better identify where problems
occurred and how interventions could be improved, which resulted in more effective interventions.

These examples highlight that theoretically informed interventions are not enough in and of themselves
to produce effective tools. Ignoring the systems to which human activities belong can lead to the development of
costly and ineffective interventions. These examples also illustrate how our theoretical leanings can prevent us
from paying attention to important and potentially confounding factors, leading to a biased understanding of the
phenomenon under study and the needs of those we are designing for. For example, for many years, HCI believed
that all users needed was basic usability, systems that simplified tasks, so most of their studies paid attention to
designing intuitive systems and largely ignored how those systems made users feel and interact with others. They
also did not carefully consider the impacts of these systems on organizations and society in the long-term.

Our field shares many similarities with HCI and health medicine in that we are designing tools, cognitive
and technological, with the goal of improving how students accomplish tasks and learn about complex
phenomenon. For this reason, we should consider the potential problems that may result from our own theoretical
leanings. One of the biggest practical issues that can result from ignoring multilevel systemic models is lacking
professional development and biased measurements of learning outcomes.
A key concern that should be central to our field, is the relative lack of impact we have had on classrooms and schools. The work on Design Based Implementation Research (DBIR) seeks to address some of the issues that emerge from traditional research paradigms that have been central to our field. While there is great challenge in creating a CSCL activity to engage learners in lab settings, and great value in understanding the nuances of the different features that we can design, approaching design from an implementation standpoint requires a different stance. The learning opportunities we have created and tested in controlled settings, are rarely suitable for use in more typical learning contexts (Penuel, Fishman, Cheng, & Sabelli, 2011). Taking into account the multiple levels of the learning ecology within which they could be used during design is necessary if the tools and activities are to be useful in contexts outside the immediate research venue.

When conducting CSCL research in classrooms, it is hard to ignore the ecological nature of learning. Students are individuals, who make sense of information in individual activities and when interacting with others, and their learning can be assessed in individual post-tests after a collaboration. However, knowledge building does happen within groups, and groups of students talk to other students within the classroom during collaborative activities. The teacher intervenes in groups, working with individual students, groups of students, or deciding to discuss an issue with the whole class (e.g. Mercier, 2016; Webb, 2009). Teachers also use consolidation activities during or at the end of tasks, asking group members to share their progress or final answers, which can serve as both an assessment exercise and a learning opportunity for other students (Joyce-Gibbons, 2017; Kaendler, et al., 2015). Finally, students learn through participation with the wider community, bringing their prior knowledge and experiences to the classroom (Moll et al., 2009), and with and through the tools and technologies of the culture within which they are situated (Barron, Walter, Martin, & Schatz, 2010). If we focus our design activities on one of these levels, and only look for learning at that level, we are likely to provide important insight into the nature of collaborative learning at that level on analysis, but be unable to determine how these changes impact the larger system. Without this understanding, we may be unable to generalize positive effects or prevent the development of unintended negative consequences. In no way do we argue that the field should abandon the careful analysis of collaborative learning, but, alone, it is not sufficient for the field to have a significant impact on practice.

Moving forward: Adopting a cognitive ecological framework in CSCL

The idea that learning is a multi-level systemic phenomenon is not new. Many scholars before us have made this claim (Barron et al., 2009; Cole & Engestrom, 1993; Jonassen & Rohrer-Murphy, 1999). Prior to the information age, Greeno, Collins, and Resnick (1996) discussed the relationship between underlying theories of cognition and learning and called for the need to rethink how we use them to inform educational research. They argued that questions about theoretical frameworks should not be limited to whether or not use of theory is coherent and leads to predictions; we need to start asking whether how we are using these theories is working. Current trends with educational design and technology use necessitate that we take up this call and start asking whether our existing theoretical frameworks are sufficient to inform practice, design, and the identification of tradeoffs associated with different instructional practices.

Our understanding of the different frameworks and the growing complexity of computer supported collaborative contexts suggests it is necessary for CSCL to begin moving towards an ecological framework of cognition. Doing so would help to relieve tensions in our field and better inform how we examine and design for collaborative learning. We could more fully investigate the trade-offs associated with different technologies and interventions and enhance our evolving understanding of how learning happens in social contexts. For example, studies that measure the impacts of technology on individual learning outcomes could also examine the impacts of the technology on social discourse. Studies that examine group learning could also examine impacts on individual learning processes and changes in the community. No one researcher would have to investigate the entire system, but they could partner with those that specialize in a complementary form of cognition, learning, or social interaction or identity development.

Adopting an ecological framework does not mean that we need to abandon previous theories of learning, but rather that we synthesize and extend them. It means we must bridge our epistemic silos, and treat knowledge and cognition as occurring at multiple levels simultaneously. If we begin to account for the notion that knowledge and cognition exist at multiple levels and are affected by multiple interconnected factors, we can move towards a deeper understanding of learning.

The implications of taking an ecological approach are not trivial. We may have to revisit how we design and report research studies. When designing studies, we need to ask questions that explore learning at multiple levels of scale and distribute strands of analysis among experts with different theoretical underpinnings in order to examine the impacts of technological interventions on a wider aspect of the ecological system. It would require us to modify how we report studies so that we acknowledge the limitations and possible unintended consequences when not accounting for learning and interaction multiple levels of scale. It would also require that we describe...
and consider the learning context in detail, considering the implications of the particular context and how that influences the generalizability of claims. The evolution of our theoretical frameworks may facilitate the development of deeper understanding of learning, the creation of more generalizable interventions, and growing impact on policy and practice.

Endnotes
(1) We note the relevance of this argument to the wider field of the Learning Sciences; due to our focus on design-based implementation research, social processes, and the limitations of the format, we have limited our paper to focus on CSCL.

References


Youth, Learning and Social Media in K-12 Education: The State of the Field

Emilia Askari, Diana Brandon, Sarah Galvin, and Christine Greenhow (1)
e.s.askari@gmail.com, dlbrandon17@gmail.com, sgalvin119@gmail.com, greenhow@msu.edu
Michigan State University

Abstract: Prior reviews of social media in education found that most early research focused on college students. This review analyzes emerging research on social media with younger learners. Using PRISMA standards, we considered 1321 articles, selecting 48 reviewed here. We find that research on social media in K-12 learning is increasing rapidly, yet largely failing to study impacts on learning. Research is taking place across the globe on a variety of social media platforms, but the majority continues to look at the oldest K-12 learners. Few studies use control groups or comprehensively use analytic data collected by social media platforms. We highlight opportunities for scaling K-12 learning with social media globally and call for educators to engage more with social media developers and owners. With this overview, we offer an agenda for future learning sciences research, policy, practice and commercial development in this growing field.

Introduction
Educational researchers and professionals have theorized the potential of social media to transform learning in and out of classrooms (Greenhow & Askari, 2017; Manca & Ranieri, 2012). Educational design initiatives have shown that engineered Facebook applications can support knowledge building and discussion in informal (Greenhow, Gibbons and Menzer, 2015) and formal higher education settings (Tsovaltzi et al, 2014). However, we know very little about K-12 students’ processes, perceptions and outcomes related to learning with these technologies. Unlike studies in the learning sciences that have focused on isolated technology-based learning environments, open social media spaces offer expanded vistas from which to study cultures of learning and teaching. For instance, online social networks introduce tools, people, and materials to school culture that could help to break up established routines and assist students in getting feedback on their performances (Greenhow, 2009). However, the field has struggled to understand the new cultures of learning arising in such “open environments” and connect them with educational practices.

In spite of their merits, social media are still largely perceived in the public discourse as platforms for social interaction devoid of learning (Whiting & Williams, 2013). Educational policies worldwide restrict their use, some even banning social media altogether, while others are vague and unhelpful. We propose a different approach – evidence based – focusing on the learner as a key factor in the effective use of social media-enabled learning environments. In this novel synthesis of over a decade of educational research, we present the state-of-field concerned with youth, learning and social media (2007-2017). This systematic literature review aims to provoke conversation on how learning scientists can advance timely, usable research that informs policy debates surrounding social media in K-12 education and the design of learning environments.

Social media, learning and youth
Social media, like social network sites (e.g., Facebook, Ning) and microblogging services (e.g., Twitter) include: 1) uniquely identifiable profiles that consist of user-supplied content and/or system-provided data; 2) (semi-)public display of connections that can be traversed by others; and 3) features that allow users to consume, produce, and/or interact with user-generated content provided by their connections (Ellison & Boyd, 2013, p. 7). Recent usage statistics in the U.S. indicate that Facebook dominates teenager’s media practices; in 2014, 78% of youth aged 12-17 reported using Facebook (Elliott 2014). Learning applications and pedagogy that build on these routines may help bridge formal and informal learning by situating social learning opportunities within students’ everyday online contexts and appropriating peer interactions on both curricular and extracurricular topics. The theoretical foundation of research on social media, learning and youth is unsettled, with scholars citing a wide range of explanations for how and why the learning occurs. Social learning theory as articulated by Vygotsky (1978) is one popular foundation of research in this emerging field. So is Gee’s (2015)’s affinity space theory, Jenkins’ (2009) writings on participatory cultures, Wegner (1998)’s ideas about communities of practice and Seaman (2008)’s work on collectivism. Several studies included in this review are based on Ellison, Steinfield and Lampe (2007)’s work on social capital in digital spaces as well as Goffman (1959)’s theories about identity
formation. Scholars suggest that social media can facilitate learners’ collaborative knowledge construction; accessing specialized just-in-time information, contributing to the hybridization of expertise; relational development and peer/alumni support especially in times of transition; academic help-seeking; social and civic benefits; and the blurring of boundaries between learning spaces, social spaces and leisure spaces (Manca & Ranieri 2013). Other scholars assert that these environments should be used for social reasons not for formal or informal learning (Crook, 2011; Selwyn, 2010).

To date, reviews of the educational research on learning and teaching with social media provide little guidance to teachers or designers on best practices for applying these technologies in K-12 education. Published literature reviews have focused mainly on the perceptions and experiences of college students (Aydin 2012; Manca & Ranieri 2013) and higher education faculty (Forkosh-Baruch & Hershovitz 2012). Moreover, the benefits of appropriating these technologies into learning contexts are contested in this research. Some studies suggest their affordances for interaction, collaboration, information and resource sharing (Maxman & Usluel 2010); encouraging participation and critical thinking (Mason & Rennie 2006; Ajjan & Hartshorne 2008); and increased peer support and communication about course content and assessment (DirVall & Kirwin 2012). Other researchers warned against exploiting social network sites for learning. Kirschner and Karpinski (2010), for instance, found that time spent on Facebook negatively affected college students’ grades. The purpose of this systematic literature review is to assess the current state of educational research on K-12 youth, learning, and social media (2007-2017) to set a programmatic agenda for the future of learning sciences research in this area.

**Connection to ICLS 2018 and partner conference themes**

This literature review connects to the ICLS 2018 Conference theme by contributing to our understanding of the complexity of K-12 learners’ processes, perceptions and concerns when social media, and designed applications, are embedded in different educational and cultural contexts. This research synthesis will also contribute to the knowledge base on how learning may be facilitated with existing social media platforms and designed applications.

Furthermore, this paper connects to themes of the Festival of Learning partner conferences: AEI (concerned with ubiquitous learning environments, informal learning, and social networks), Learning@Scale (concerned with opportunities to scale learning), and the development of commercial educational technology by advancing knowledge of the state-of-the field of youth, learning and social media research to foreground opportunities to design and study educational applications that run on global social networks, potentially leading to commercialization. Acknowledging that the most popular social media are owned by commercial companies, and are monetizing information collected from users, we offer ideas about how educators might engage more powerfully with these companies to shape the future of these important communication and education platforms.

**Methods**

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Standards guided the methodology for this literature review (Moher, et al., 2009). In consultation with a library science expert, the team identified four prominent databases for educational research: Education Full Text, ERIC, Scopus, and Web of Science. We performed database searches using the following terms: “(K-12 Teaching OR teacher training OR professional development OR elementary school or middle school or high school) AND (social media or social network sites or social networking sites or Facebook or Twitter or microblogging or Pinterest or Instagram) NOT (higher education)” finding 1268 articles. Note: this paper will address themes in the articles concerned with K-12 learners only and not themes in identified articles related to K-12 teachers and preserve teachers as this coding is still ongoing. The team reviewed each title and abstract, evaluating the relevance of each record. Criteria for an article’s inclusion were: a focus on K-12 learning (e.g., students’ learning in- or outside of school) and an inclusion of social media within our parameters (e.g. social networks, microblogs, collaboration tools, etc.). Articles were excluded if they were outside of K-12 education (e.g. nursing school, undergraduate education, etc.).

Furthermore, 100 journals publishing education or educational technology research were identified based on previous reviews (Manca & Ranieri, 2013). Any journals not covered in the database search underwent a manual table of contents search for the years available: Journal of Literacy and Technology (Dec. 2007 - Present), Journal of Educational Technology Development and Exchange (2011 - Present), and International Journal of Online Pedagogy and Course Design (2011 - Present). Two journals were unavailable online and, therefore, excluded: International Journal of Continuing Education and Lifelong Learning and International Journal of Technology in Teaching and Learning. We also scanned the bibliographies of articles deemed relevant. Through this process, we identified an additional 53 articles from other sources; we removed 8 duplicates, leaving a total of 93 articles.
Those 93 articles were then rated by three independent researchers based on the abstract and as needed, the full text. Any discrepancies between coders were discussed and resolved. This process lead to the exclusion of 42 articles deemed not relevant, leaving a total of 50 relevant articles. Of these 50 articles, two were excluded because they could not be accessed. During the process, each researcher kept notes about search and relevance decisions, in keeping with the PRISMA standards.

The 48 full-text articles were then read and coded by the team and analyzed for themes. Based on previous literature reviews (e.g. Greenhow & Askari, 2017; Manda & Ranieri, 2013; 2016) five categories were identified as important to note in analysis: country context for the study; social media platform; age and/or grade level of participants; methodology; type of learning: formal, informal or other. Further, coders looked to the data to derive additional themes, allowing the data to suggest categories outside those established in the literature.

Findings
The overarching conclusion of this review is that research into uses of social media in K-12 education is rapidly increasing. Where earlier reviews of social media in learning found only a handful of studies with participants in the K-12 age range, this review identified 48 studies published in peer-reviewed journals, most within the last five years. Yet, within that sample, participants still skewed towards the older end of our age spectrum. Only six of the 48 studies in this review had elementary school aged participants (5-11 years old). Seventeen had participants of middle school age (12-15 years old) while 36 had participants of high school age (16-18 years old).

While Facebook remains the most prolifically studied social media platform (22 studies or 46%), the breadth of the literature is expanding. The next largest percentage of studies reflects this trend: ten (or 21%) were nonspecific in their research, collecting data on social networking sites in general rather than limiting findings to a particular platform. Sixteen different types of social media were represented in the articles; one was designed by Mallan, Singh, and Giardina (2010) specifically for their learners.

The diversity in social media studied is likely a result of the global interest in learning with social media. With 21 different countries represented in 48 articles, it is clear that the impacts of social media on K-12 learning is a worldwide concern. Furthermore, no single country has a monopoly on this the work; the country where the most studies (10) were conducted, the United States, still accounts for less than 20% of the places researched. England, with eight studies, and China, with five, were the next most frequently represented countries in our dataset. Only one study (Lantz-Andersson, Vigmo & Bowen, 2013) compared social media use by K-12 students in different countries (i.e., in Colombia, Finland, Sweden and Taiwan).

Roughly equal numbers of studies focused on informal (19 studies) and formal (22 studies) learning contexts. Although definitions of informal and formal learning are contested in the research literature (Greenhow & Lewin, 2015), informal learning here refers to unconscious personal development (e.g. identity work) or conscious, but voluntary learning that is not a requirement of the classroom curriculum (Greenhow & Lewin, 2015). Formal learning may be driven by teaching objectives or curricula associated with K-12 schooling. Four studies straddled informal and informal learning, while three studies focused on describing student practices on social media that connected to learning without anchoring the study to any specific learning environment. For instance, Cook (2012) engaged hundreds of English high school students in focus group discussions of the differences in their social media practices in school and outside of school. Ma and Chan (2014) surveyed 299 Hong Kong high school students about their motivations for sharing content via social media, and Lee (2015) surveyed 720 middle and elementary students in Korea about negative behaviors on social media such as using improperly using digital content without the owner’s consent.

In examining topics taught in classrooms where formal learning studies were conducted, the most common subject addressed was language arts, including writing and digital media, which served as the focus for 13 studies, explored under the Media Literacy heading. Six studies focused on teaching study skills or collaboration within a formal classroom where other subjects of learning were not specified (Andersson, Hatakka, Grönlund & Wiklund, 2014; Fewkes & McCabe, 2012; Kio, 2016; Rosen, Carrier & Cheever, 2013; Vasbø, Silseth & Erstad, 2014; Vilca & Vallejos, 2015). Five studies analyzed social media use in science- or technology design-related classrooms (Huang, Wu, She & Lin, 2014; Lai et al., 2015; Reynolds & Chiu, 2016; Veira, Leacock & Warrican, 2014; Won, Evans, Carey, & Schnittka, 2015). As an example of the studies connecting social media learning to science learning in k12, Won, Evans, Carey, and Schnittka (2015) surveyed 44 students at a Virginia middle school, documenting the utility of Edmodo for supporting collaborative design process in an afterschool program, Studio STEM. Three studies took place in classrooms where English as a Second Language was taught (Lantz-Andersson, Vigmo & Bowen, 2013; Sun et al., 2017; Vikneswaran & Krish, 2016).

A strong preference for surveys emerged in methodology across the studies. A total of 26 studies (54.2%) used a survey. Thirteen studies (27%) used only surveys as their source of data, with eight focusing on informal learning and five divided between formal and informal learning.
The number of participants in the 48 studies in this review ranged from one, for a case study analyzing one English student’s Bebo profile page (Dowdall, 2009), to 2611, for a survey of social media use by high school students in England and Wales (Luckin, et al., 2009). Fourteen studies, nearly 30% of the articles in this review, had 50 participants or fewer including five studies with ten participants or fewer. Twenty-seven studies, (56%) had more than 100 participants, including six with 500-1000 participants and three with more than 1,000 participants.

Twenty studies (or 41.7%) included some quantitative analysis of social media activity such as number of posts. An additional eight studies included observation of social media or implementation of social media as part of the procedures, making a total of 30 studies (62.5%) that directly engaged with social media. The other studies asked participants to describe experiences using social media to support student learning.

Only five studies (or 10.4%) used a control and treatment group design (Baris & Tosun, 2013; Cano, 2012; Ch., Mahmood, & Rasood, 2016; Lai et al., 2015; Sun et al., 2017).

Impact on learning outcomes
One of the most startling trends identified was the dearth of attempts to measure social media’s impact on student learning. Only nine studies, including the five control and treatment design studies mentioned above, presented such findings. Baris and Tosun (2013); Cano (2012); Huang, Wu, She, and Win (2014); Lai et al. (2015); and Reynolds and Chiu (2016) all found content-specific grade improvements because of social media implementation. Sun et al. (2017)’s work with English as foreign language students found that social media is also effective in improving English language skills. In an interesting disagreement, Ch., Mahmood, and Rasood (2016) and Khan, Wohn, and Ellison (2014) found correlations between social media use and higher grades, while Rosen, Carrier, and Cheever (2013) found that social media use led to lower overall GPAs. Although the majority of studies reported success with social media in K-12 classrooms regardless of acknowledged challenges, these findings, in most studies, are not supported by data on changes in student learning connected to social media.

Social media and literacy
Throughout the literature, social media is seen as creating both a new path toward learning traditional language arts skills as well as a path toward gaining new digital skills – media literacy – that prepare students for the 21st century. Karal, Kocok, and Cakir (2015), for example, found that Facebook helped improve 30 Turkish high school students’ writing skills, analyzing how their interactions and relationships between students themselves and between the teacher and students were essential to their growth. Li and Wu (2017) used a survey and focus group to study how 1,039 high school students in China practiced reading on WeChat, investigating their motivations through a cultural and social lens. Cano (2012) found that 280 high school students performed better on Spanish assessments after integrating Twitter into the curriculum.

Beyond bringing new perspectives and methods to traditional literacy skill sets, Reynolds and Chiu (2016)’s work with 242 middle and high school students suggests media literacy may help address the digital divide. The study’s data comparing students’ computer skills growth with the education levels of their parents showed that students whose parents had lower education levels improved more than those whose parents had higher levels of education after the implementation of the game-design social media, Globaloria. Although Lu, Hao, and Jing (2016) did not find a similar correlation between parent education level and student social media use in their survey of 186 Hong Kong high school students, these two studies may still be pointing towards the same finding: the potential for media literacy learning is not dependent on parent education level; media literacy programs in schools can create equal competence in students from varying backgrounds.

Luckin et al. (2009) explains that social media has the potential to be an asset in students’ learning, but that a lack of the media skills required to complete complex tasks online is preventing effective application. Based on their case study of three 5th grade girls’ usage of Reading Revolution, Lindstrom and Niederhauser (2016, p.116) build on the idea that students need a new skill set to be academically productive online: media literacy is “a new cultural form that builds on established rules and conventions, and is neither right nor wrong, but simply constitutes a new way to communicate.”

Peer-to-peer learning and hybridized expertise
Social learning as an aspect of media literacy was highlighted in 13 studies, underlining the affordance of social media to continue peer-to-peer interaction and learning beyond specified school times and places. As one of these studies – Khan, Wohn and Ellison (2014, p. 139) – points out, “Evidence shows that the cognitive processes for deep learning and information retention processes occur in dialogs.” In traditional educational setting, this discussion occurs in a classroom. Social media extends those conversations, which numerous studies found to be critical to learning, and which also mimic the kind of learning that occurs in many work situations the students
will encounter later in life. Khan, Wohn and Ellison (2014) collected survey responses from 690 US high school students, finding that Facebook enabled students to collaborate effectively on learning activities, including homework and resource-seeking.

**Civic learning and engagement with the world outside of school**

Three studies highlight the affordance of social media to connect learners with the world beyond the classroom. Beach and Doerr-Stevens (2011) used Ning as a platform for capturing the argumentative exchanges of 30 students in a U.S. high school participating in a virtual debate about school policies. The authors concluded that social media can be beneficial for forefronting competing opinions on an issue and encouraging response towards building collective understanding if not always agreement. Huang, Wu, She and Lin (2014) analyzed the Facebook posts of 83 Taiwanese high school students engaged in group conversation about science news as a tool for collectively developing understanding of the nature of science. Charitonos, Blake, Scanlon and Jones (2012) used Twitter to help 29 British 13- and 14-year-olds share ideas about artifacts explored on a field trip to The Museum of London.

**Steps towards bringing social media learning to scale**

Four studies reach toward developing tools that might be used to support learning by young people via social media at scale, envisioning replication in educational contexts around the world. Oussalah, Escallier and Daher (2016) is among them. The authors created an automated system for analyzing Twitter posts by 133 high school students in England. The system explored ways to assess the grammar used in these posts, and was supplemented by surveys asking students about their perception of the experiment. While the authors conclude that abbreviations and slang that are commonly used on Twitter create challenging obstacles for automated analysis of grammar, this study takes a step toward potential future development of similar tools that might be deployed with large numbers of learners.

In addition, the three studies looking at using social media to learn English as a Second Language offer potential guidance for developers thinking about using social media at global scale in K-12 learning. Given a British Council (2014) report declaring English to be the dominant global language with 1.5 billion learners studying English as a Second Language, and noting that “there appears to be a fast-moving worldwide shift, in non-anglophone countries, from English being taught as a foreign language (EFL) to English being the medium of instruction (EMI) for academic subjects such as science, mathematics, geography and medicine.” (Dearden, 2015, p. 4), these studies suggest how social media might be used to address the planet-wide market for English learning. Unlike many other subjects, English grammar and usage may be less negotiable and influenced local curricular preferences than some other subjects, and thus perhaps more suitable for learning at scale. Yet, we acknowledge that all languages evolve, and people learning English as a Second Language are participating in that process (Scockt & Toffoli, 2012). Studies with college students have found that social media is seen as a good place to learn English (Kabilan, Ahmad & Abidin, 2010; Yunus, Salehi, & Chenzi, 2012) -- a concept that could reach an even wider audience of informal English learners (Scockt & Toffoli, 2012). Turning to the k12 studies in this review focused on learning English as a Second Language, Vikneswaran and Krish (2016) examines the motivations of ten Chinese-speaking high school students in Malaysia who are asked by their new school to use Facebook to help them learn English. As Vikneswaran and Krish (2016, p.298) observes, "writing on social media platforms is generally informal, hence making it easier for users who are not very proficient in the English language to contribute their thoughts and ideas” without expectation of judgement, forefronting more intrinsic motivations related to learning and communication with peers. Lantz-Andersson, Vigeo and Bowen (2013) analyzed a closed Facebook group created for use by a total of 60 high school students studying English as a second language in Colombia, Finland, Sweden and Taiwan. The authors found promise for using Facebook as a medium for language learning across cultural context, noting, however, that such efforts must be “deliberately and dynamically negotiated by educators and students to form a new language-learning space with its own possibilities and constraints” (Lantz-Andersson, Vigeo & Bowen, 2013, p. 310). Sun et al. (2017) analyzed how 72 Chinese first graders studying English used Chinese-based social media platform focused on sharing photos and audio files, Papa. The authors found that students who communicated with each other in English by sharing recordings improved significantly more than a control group. This shows potential for using Papa and similar applications at scale, in other settings -- both formal and informal -- where English is not the native language.

**Discussion and implications**

Our findings suggest opportunities for research, facilitation of learning, education policy, and commercial development of educational technology, as discussed in these next sections.
Agenda for research

The field of research on social media in K-12 education is still young. We see numerous opportunities for future studies in with this population – particularly focusing on social media’s impact on learning outcomes. Additionally, more studies are needed with elementary school-aged participants, who are the least-studied age group in this review. Further, we hope to see more studies using data related to social media posts actually made by students, as opposed to more studies asking students how they use social media. Nevertheless, we see this data on actual posts as one component of potentially rich ethnographies of student social media use, which would be bolstered by student interviews exploring motivations and impacts related to social media usage. Noting that we reviewed only research published in English, yet observing that many of the studies in this review were conducted in countries where English is not the native language, we call for reviews similar to ours analyzing research printed in other languages – and ideally sharing that analysis with the English-speaking research community. Further, we see a need for more comparative studies looking at use social media use in K-12 learning in different language and culture contexts.

Facilitation of learning: implications for teaching practices

As many studies in this review point out, facilitating learning with social media in K-12 settings presents teaching challenges. Therefore, we offer three practical suggestions for K-12 teaching context. First, use social media to create a sense of community and a space to build relationships. Social media enables students to create a comfortable, supportive environment, where learning risks and identity development can be explored without fear of judgement (Eamer, Hughes & Morrison, 2014; Erjavec, 2013; Vikneswaran & Kirsh, 2016). Encouraging students to use social media as a way of blending identity and academic work allows students to have an authentic experience and reveals details of students’ learning practices that can inform teachers’ practices (Lindstrom & Niederhauser, 2016; Vasbo, Silseth & Erstad, 2014). Forkosh-Baruch, Hershkovitz, and Ang (2015) looked at differences between teachers and students who were and were not willing to connect on Facebook, and this theme of understanding the significance of the changing relationships that social media incurs is also emphasized in the work of Callaghan and Bower (2012) and Yang, Crook, and O’Malley (2014).

Second, teach students how to use social media for more than the transfer of information; integrate creative projects in which students are producing content. Given more training in advanced skills, social media does not need to be a distraction for students in the classroom; it can be a vehicle for discovery, constructivist learning, and imaginative thinking (Charitonos, Blake, Scanlon, & Jones, 2012; Fewkes & McCabe, 2012; Khan, Wohn, & Ellison, 2014; Lu, Hao, & Jing, 2016; Luckin et al, 2009; Ranieri & Bruni, 2014). Mallan, Singh, and Giardina (2010) called for more student involvement in teachers’ quests to understand the potential of technology in the classroom; youth adapt technology to their needs just as quickly as technology improves, and they can be a resource for discovering useful social media tools.

Third, maintain a strong teacher presence and accepting social media as an integral part of students’ lives. Andersson, Hatakka, Grönlund, and Wiklund (2014) proclaim that teachers need to “reclaim the students,” and rather than thinking of social media as a “beast,” teachers should embrace more structure and stricter management (p. 49, 47). Social media cannot be ignored in the classroom; it is a “hidden curriculum” fully integrated in youths’ learning whether teachers accept it or not (p. 22). Rosen, Carrier, and Cheever (2013) point to the impossibility of expecting students to disengage from their digital lives during class; students will always check their phones, or worry about checking their phones; instead of fighting a losing battle, consider options that ease students’ multitasking compulsion without detracting from learning such as offering “technology breaks” (p. 956). As Casey and Evans (2011) summarize, “Teachers cannot take this approach in fear of chaos and disorder; they must find innovative ways to construct disorder and flow with chaos and build resilience to the traditional training that instinctively drives them to take control” (emphasis added).

Implications for education policy

Our results suggest the need for shifts in education policy at all levels – school, district or region, state and national. If we acknowledge that K-12-aged students are growing up with social media as large component of global culture, then we need educational policies at all levels that support teachers’ use of these important communication platforms for learning. School-level policies that block access to social media sites should be reexamined. Curriculums that are silent on use of social media should be amended to include suggestions on how to use social media effectively. In addition, teachers should be offered guidance on how to manage their professional social media accounts in ways that will model good digital citizenship. When teachers observe misbehavior on social media, there should be policies in place to encourage resolution and turn these mis-steps into learning experiences. We note that while the tone of most studies in this review is positive, our search criteria eliminated studies focused on cyberbullying and social-media related depression, where these studies were not explicitly also connected to
learning. Nevertheless, some studies in this review noted learning-related downsides to social media use, such as inattention and lower grades. In this context, we wonder why educators are ceding control of data about how young people are using social media in education to the for-profit companies that create social networking platforms. The data collected by Facebook and other popular social networking sites on users – including our youngest students – is analyzed by Facebook and its competitors and sold to private data brokers for marketing purposes (Ramirez, Brill, Ohlhausen, Wright & McSweeny, 2014). As educators, we ask: why should we allow these companies to make money from our students’ use of these important platforms of communication and literacy development? Why isn’t it easy for researchers to get data about uses of social media in educational contexts? Since the Terms of Service created by these companies require users to give up much of the data about their use of these platforms for marketing purposes, why shouldn’t these companies be compelled to share the data for research purposes as well? We note that Facebook Research partnered with researchers on several oft-cited studies with college student participants (Ellison, Steinfield & Lampe, 2007; Ellison, Steinfield & Lampe, 2011). That work and other studies conducted with college students should be extended to younger students, especially since developmental theory suggests that early adolescence is a time when friendships and related social capital are intensely important (Berndt, 1982). We are eager to see collaborations between social media companies and researchers working in K-12 education contexts. Under U.S. law, major social media companies are allowed to design for users who are 13 or older, so there should be no legal restraints to conducting large-scale research on social media with US students in middle school and high school. We call for policies at all levels of education pushing for greater access for researchers to data collected by social media platforms that may be owned by large for-profit corporations, yet are increasingly used as essential tools in public education. By blocking researcher access to this data, school policies are not protecting students’ privacy. Instead, such policies merely ensure that only the for-profit companies that own the major social media platforms are seeing this data.

Implications for the commercial development of educational technology
Companies such as Facebook and Twitter are powerful enough to compare their executives to global political leaders. These companies profit from selling the data they collect about all their users, including students and teachers, to companies that want these users to buy commercial products and services. Social media companies, notably Facebook, have been widely criticized for designing algorithms to manipulate user’s emotions to provoke engagement with paid content. Yet it is possible to envision newer, innovative platforms replacing or remaking these iconic social media brands. In this discussion of future developments of social media in education, we draw attention to the ideas of the high school student participants studied by Bowler, Knobel and Mattern (2015, p. 1274), who suggested that in the future social media developers should “design for hesitation, design for consequence, design for empathy, design for personal empowerment, design for fear, design for attention, and design for control and suppression.” For example, social media could employ more “undo” buttons, giving users the ability to edit or even erase social media activity over a longer period of time. Social media could use algorithms to connect people who might be impacted by the topic of a post with people who are discussing that topic from a distance on social media. We think such features would enhance learning opportunities with social media in K-12. We hope that the public-interest perspectives of K-12 students and teachers -- and the researchers who study them -- will help develop the next generation of social media.

Endnotes
(1) Authors contributed equally and are listed alphabetically

Abbreviated References
Due to space constraints please find the full Reference List here: http://bit.ly/youthlearningsocialmedia


Defining and Assessing Risk Analysis: The Key to Strategic Iteration in Real-World Problem Solving

Spencer E. Carlson, Leesha V. Maliakal, Daniel G. Rees Lewis, Jamie Gorson, Elizabeth M. Gerber, and Matthew W. Easterday

s.c@u.northwestern.edu, lmaliakal@u.northwestern.edu, daniel.rees.lewis@u.northwestern.edu, jgorson@u.northwestern.edu, egerber@northwestern.edu, easterday@northwestern.edu

Northwestern University

Abstract: Across domains from science to civics, experts plan to solve real-world problems iteratively. Despite the importance of strategic iteration, we lack precise understandings of effective iterative planning and novice challenges, making it difficult to assess formatively and therefore to teach. We conducted design-based research to understand iterative planning and design assessment tools in a full-time 6-week program where undergraduate teams worked on social impact design problems. We found that iterative planning requires a process of risk analysis: detecting risks in the problem space, prioritizing those risks, and setting goals to reduce them. Novices struggled with each step of risk analysis, so they did not plan iterations strategically. We designed assessment tools that surface students’ thinking about risk analysis and support instructors to notice common challenges. We contribute detailed understandings of iterative planning and novice challenges as well as tools that can be adapted to assess real-world problem solving across domains.

Keywords: problem solving, design learning, iteration, planning, assessment, risk analysis

Iterative planning to solve real-world problems

Most real-world problems cannot be solved in one try—they require iteration: the practice of testing and revising ideas to continually improve one’s understanding of the problem and solution (Adams, Turns, & Atman, 2003; Wynn & Eckert, 2017). Scientists iterate on research questions to refine their contributions, engineers iterate on technical solutions to satisfy practical needs, and doctors iterate on diagnoses to prescribe more effective courses of treatment. Iterative planning—the process of strategically planning to iterate—is thus critical to teach when preparing students for practice in problem solving disciplines.

In iterative planning, professionals choose activities that enable them to continually improve both their understanding of the problem and their solution (Adams et al., 2003; Crismond & Adams, 2012). For example, a scientist might iteratively improve their argument in a research paper by planning to collect, analyze, and write up additional data. This approach to planning contrasts with teaching students to choose activities (like data collection, analysis, and writing) based on a predetermined project schedule. Unfortunately, novices struggle to iterate strategically (Ahmed, Wallace, & Blessing, 2003; Crismond & Adams, 2012).

Need for tools to assess iterative planning

If we want to teach iterative planning, we need formative assessments of iterative planning to help teachers decide where to focus instruction. To assess iterative planning reliably, assessors need to make consistent decisions about what information to seek when assessing performance, and how to assess it. Assessment tools could support consistent decision making when assessing iterative planning.

Previous work suggests assessing problem-solving by using performance rubrics and assessing argumentation. For example, it is common to assess the arguments students construct to justify their problem-solving activities. However, simply assessing argumentation does not directly assess iterative planning.
solving (Cho & Jonassen, 2002; Shin, Jonassen, & McGee, 2003). Highly-ill structured problems like design problems underlie professional practice in most disciplines (Jonassen, 2010). In these problems, assessing the solution itself is difficult because the criteria for evaluating design solutions are highly subjective, change with one’s understanding of the problem, and are not fully known until the end of the design process (Jonassen & Hung, 2015). So, if we want to assess this type of problem solving, we cannot simply check whether students found the right solution. Rather, we need to assess how reasonable a proposed solution is, given one’s current understanding of the problem. However, assessing the quality of students’ arguments is an ambiguous task without clearly defined criteria for argument quality, which makes reliable assessment more difficult. To remedy this, Jonassen (2010) suggests using argumentation rubrics that define criteria for assessing students’ problem solving performance.

Given that iterative planning is a key process within problem solving, we hypothesize that argumentation rubrics can also be used to assess performance in iterative planning. Like in solving design problems, there is no single correct solution in iterative planning. There are poorly-justified plans, and there are better-justified plans. If we can create rubrics that define poorly justified versus well justified plans, these rubrics might support reliable assessment of iterative planning performance. However, prior research does not provide specific guidance about tools and practices for constructing rubrics to assess iterative planning because iterative planning is currently an ill-defined task. And so we ask: how do experts reason through iterative planning and where do novices struggle? By answering this question, we can construct rubrics that effectively assess whether novices exhibit the desired performance.

Need for novice and expert models of iterative planning
Unfortunately, the literature tells us little about how experts (i.e., professionals) and novices (i.e., students) reason about iterative planning; that is, we lack expert and novice models of iterative planning that are detailed enough to create assessment rubrics. Experts solve design problems by iterating: refining their understanding of the problem and solution as new information emerges during problem solving (Adams et al., 2003; Adelson & Soloway, 1985; Atman et al., 2007; Guindon, 1990). However, we know little about how experts reason through the iterative planning process. Nor do we know what novices find difficult in that process. For example, a recent literature review on expert and novice design practices noted the importance of iteration (Crismond & Adams, 2012) but did not identify how experts carry out iterative planning. The same review established that novice designers undervalue and underutilize iteration, but did not identify why novices struggle to plan iteratively. We need to define how expert designers reason through iterative planning—and where in this reasoning process novices commonly struggle—to create assessments of iterative planning performance.

Research questions
The goal of this project was to develop reliable tools for assessing iterative planning performance, and corresponding expert and novice models which inform those tools. Specifically, we asked: (a) How do expert designers approach iterative planning? (b) What do novice designers typically struggle with when they attempt iterative planning? and (c) How can we assess design students’ performance in iterative planning?

In this study we conducted 6 1-week iterations of data collection, analysis, and redesign to create tools for assessing student teams’ planning in design projects. We developed expert and novice models of iterative planning to design a rubric tool to assess planning. We also developed two tools for externalizing students’ plans: the Design Canvas (DC), the Iteration Plan (IP). The DC and the IP are complementary, poster-sized (40 in. x 60 in.) templates where students externalize the knowledge and reasoning they use to plan (for images see our companion study: Rees Lewis et al., 2018). This happens during collaborative planning in a design team, similar to professional design practice (Osterwalder & Pigneur, 2010). The rubric guides assessors to detect students’ struggles by reviewing the DC and IP.

Methods
Context and participants
We conducted the research in a 6-week extra-curricular undergraduate summer program at a university design institute. In the program, each 4-5 person project team worked with a local community partner organization to design products and services to address a given real-world challenge for the duration of the program. Challenges included: improving airport accessibility for autistic travelers, reducing air-travel-related wheelchair breakages, improving accommodations for people with dementia, increasing first responder support for youth, and reducing teen depression. Teams worked approximately 36 hours a week to create an original solution based on their
understanding of the problem. Each week, we prompted students to plan. First, students updated their DC so that it reflected their current knowledge of the problem space. Then, students constructed and represented a plan using the IP.

The study involved 21 undergraduates from 18-22 years old at a large private US university. Participants majored or double-majored in engineering (15), natural sciences (3), social sciences (3), art (3) and journalism (3), and included 4 first-, 12 second-, and 5 third-year students (57% female).

Data collection and iterative analysis
We collected 1 photograph per week of each team’s DC and IP, after the teams had planned. In total, we captured images of 30 plans across 5 teams. We also wrote field notes based on observations of team planning sessions. After each weekly planning session, we reviewed photographs and field notes to critique teams’ planning. When we made a critique that was not already represented in the rubric tool, we added it to the rubric and wrote an analytic memo explaining why we would make a different planning decision. This process helped us to articulate the normally tacit knowledge we use to think about planning decisions. In many cases, assessing the quality of the plan required additional information, such as students’ reasons for choosing a certain goal. When we found we needed additional information (beyond what we saw in the Design Canvas or Iteration Plan) to assess a plan, we revised the tools to prompt students to share that information. We also wrote an analytic memo explaining why we needed that additional information to judge the planning decision. Like with the memos we wrote to justify critiques, this enabled us to articulate elements of tacit knowledge that we use to think about planning decisions. For example, while the tools ultimately centered on a critical process of risk analysis (see Findings), the initial prototypes did not surface students’ thinking about risk. By analyzing students’ plans we recognized the centrality of risk in our own thinking, in part because it was impossible to evaluate students’ planning without understanding how they thought about risk in the problem space. This led us to begin articulating our understanding of risk in an analytic memo, and to add additional boxes to the Design Canvas prompting students to explicitly identify risks.

What authorizes us to define expertise? First, we are experienced designers with over 50 collective years of design experience in industry and academic design-based research. Second, in a companion study of this learning environment (Rees Lewis et al., 2018), we tested instruction based on our understanding of iterative planning. Students who received this instruction engaged in key design practices (e.g., iteration, interviewing users, testing ideas with stakeholders) more than students from the previous year (when we did not emphasize iterative planning). This suggests that learning iterative planning (as we define it) helps students succeed in design.

Summative analyses
At the end of the program, we conducted summative analyses to test the final versions of our models and prototypes. One researcher used the final rubric to assess students’ plans from each week. This allowed us to test the final novice model against data from each week of the program, and to aggregate our evidence for the final novice model. A second researcher used the final rubric to assess 5 of the plans we collected (17% of plans). This allowed us to calculate Cohen’s kappa (.82) between researchers to test reliability of the assessment tools.

Figure 1. Task analysis of iterative planning. We found iterative planning requires risk analysis: detecting risks in the problem space, selecting high-priority risks to address, and setting near-term goals to reduce those risks. Novices struggled throughout iterative planning; this paper focuses on challenges with risk analysis.
Findings
We found that risk analysis is a central reasoning process in strategic iterative planning. In risk analysis, designers review their mental or external representation of the problem space to assess, prioritize, and set near-term goals to reduce risks in the problem space. We found that novice designers face challenges throughout the process of risk analysis. We found that the DC and IP successfully externalized students’ reasoning about risk analysis such that we could tell where students struggled. Finally, we found we could assess students’ risk analysis reliably using the assessment rubric.

Strategic iterative planning requires Risk Analysis
As we observed teams and assessed their Design Canvases and Iteration Plans, we reflected upon their planning decisions. When we identified a decision that we would make differently, we wrote an analytic memo to (a) catalogue our critique of the team’s planning decision and (b) articulate our reasoning about why we would make a different decision. In doing so, we found that our implicit expert model of iterative planning centered on the process of risk analysis, in which designers detect risks in the problem space, prioritize these risks, and set near-term goals to reduce high-priority risks (Figure 1). We found that iterative planning also involves representing the problem space and constructing the full plan (Figure 1), but we focused on risk analysis in this study. We found risk analysis relies on several key concepts that novices must develop, including: a problem space schema (Table 1); a causal model for detecting risk (Figure 2); knowledge of the process of risk analysis (Table 2: Experts ask themselves; Table 4: What experts do); knowledge of principles justifying the process of risk analysis (Tables 2 and 4: Why it matters); and knowledge of common sources of risk (Table 2: Common sources of risk). The corresponding novice model identifies several aspects of risk analysis that students found to be particularly challenging (Table 3; Table 4).

Table 1: A problem space schema that the experts used to analyze iterative planning in social impact design

<table>
<thead>
<tr>
<th>Problem Aspects</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Partner (CP)</td>
<td>A person at a partner organization with expertise in the problem area (e.g., a relevant non-profit). The CP connects problem solvers with resources (e.g., information, access to users). The CP may also implement designers’ solution, if they find it helpful.</td>
</tr>
<tr>
<td>User Access Plan</td>
<td>How designers plan to access users (see User row) to learn about their needs.</td>
</tr>
<tr>
<td>Demoing Plan</td>
<td>How designers plan to get regular feedback from the CP.</td>
</tr>
<tr>
<td>Desired Impact</td>
<td>The social impact that designers want to have.</td>
</tr>
<tr>
<td>User</td>
<td>A persona who will use the proposed solution. Users have needs, which have 3 components: a “job” (a task users must complete), a “pain” (a challenge they face in that task), and a “gain” (the benefit they will attain if they can complete the task). By satisfying user needs, problem solvers can entice users to adopt the proposed solution (and thereby promote the Desired Impact).</td>
</tr>
<tr>
<td>Root Causes</td>
<td>An analysis of the underlying, fixable causes that explain why the user need, CP need, and desired impact are not yet satisfied.</td>
</tr>
<tr>
<td>Value Proposition (VP)</td>
<td>A proposed solution and an argument for how the proposed solution will overcome the root causes to satisfy the user need, CP need, and desired impact.</td>
</tr>
<tr>
<td>Existing Solutions</td>
<td>Designers need to account for existing solutions and argue why theirs is better.</td>
</tr>
<tr>
<td>Implementation Strategy</td>
<td>Explains who will implement the proposed solution and how. Commonly, this involves handing off the solution to the CP, who integrates it into their existing operations.</td>
</tr>
<tr>
<td>Impact</td>
<td>Evidence that designers have achieved the desired impact.</td>
</tr>
</tbody>
</table>

The central concept in risk analysis is risk—the probability that the design project fails to make impact (i.e., change the status quo in a desired way). Designers’ concept of risk derives from an implicit causal model of the factors affecting whether a design project makes impact (Figure 2). Teams increase the probability of making impact by working on the causal factors that affect impact (Figure 2). For example, for the project to make impact, the team must ensure that users adopt the designed solution, which means designing something that users value, which requires understanding user needs (see the problem space schema in Table 1 for explanations of user, community partner, and other terms). Likewise, a second causal chain focuses on understanding the needs of the community partner (CP), who also affects whether the designed solution is adopted. Furthermore, assuming the solution is adopted by users and CPs, the solution also must work as intended to achieve its desired impact. Each causal chain presents a possible risk to the impact of the project, so
design teams iteratively plan activities that help them understand and address risks of user need, partner need, and solution efficacy.

Figure 2. Experts (us) used this implicit causal model to detect risk in the problem during iterative planning.

Risk exists when any of the variables in the causal model represent unmet conditions (e.g., designers don’t know the root cause of the CP’s need, or the user doesn’t value the design solution) or unknown conditions (e.g., designers are unsure if they are right about the root cause, or they don’t know if the user values their solution). Downstream variables are closer to making impact than upstream variables, so downstream variables have a more direct causal effect on making impact and thus matter more for risk. For example, solvers can only achieve a relatively small decrease in risk by working on the Know User Need variable, because even if they understand the user need perfectly, there are many downstream variables that can interfere with making impact.

Table 2: Expert model of risk detection in social impact design

| Identify and estimate the level of risks associated with the Community Partner (CP). |
| Experts ask themselves: |
| Why it matters (connection to Figure 2: causal model of risk): |
| Common sources of risk*: |
| • We haven’t made contact with a real person at a partner organization |
| • We can’t specify the partner’s need (either as a concrete “job,” “pain,” and “gain,” or a clearly measurable social impact goal) |
| • Our ideas about the partner’s need aren’t reasonable (they conflict with data and/or common knowledge) |
| • When we cite data, we don’t specify both its content and its source |

* i.e., conditions that may threaten impact by interfering with the variables that lead to impact (shown in Figure 2)

This explains why designers cannot sufficiently reduce risk by spending all their time understanding the problem (by interviewing the partner and users, and defining desired impact): an understanding of the problem does not guarantee a working solution. For this reason, designers try to move as quickly as possible to building and testing potential solutions to see whether users and CPs actually value and adopt the solutions, and whether the solutions actually achieve measurable impact. In other words, designers plan by weighing the risks inherent in different potential plans; they must judge whether they can achieve impact most quickly by building and testing a solution, or whether spending time understanding the problem will save precious time by avoiding “building the wrong thing.” These questions rarely have a clear right answer, if ever. However, designers use the process of risk analysis to discriminate between more and less reasonable answers, based on their knowledge and experience. This involves detecting and prioritizing risks. In our study, expert designers (i.e., the authors) detected risks by analyzing aspects of the problem space, using knowledge of the process of risk analysis, the principles justifying it, and common sources of risk (Table 2). The full expert model of this knowledge does not fit within the page limit, but Table 2 provides an excerpt. The experts then prioritized risks and set goals using additional knowledge of the process and principles of risk analysis (Table 4).

Novice challenges in risk analysis
We found that students struggled throughout the process of risk analysis. Students faced challenges in risk detection (Table 3), risk prioritization, and goal setting (Table 4). In some cases, students skipped steps in the process of risk analysis (e.g., not attempting to detect risks). In others, students’ reasoning did not align across
the steps of risk analysis (e.g., setting a goal that did not address high-priority risks). In still other cases, students struggled to complete the steps of risk analysis in a way that would be useful (e.g., setting a goal that is too vague to guide the rest of planning). Students also struggled to identify all of the risks we noticed in their problems, and they often struggled to make reasonable estimates of risk level.

Table 3: Novice model of risk detection

<table>
<thead>
<tr>
<th>Problem Aspects</th>
<th>Don’t attempt to detect risks</th>
<th>Don’t both identify risks and estimate the level of risk.</th>
<th>“Risks” identified are not truly relevant.</th>
<th>Did not identify certain salient risks.</th>
<th>Estimated level of risk is unreasonable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Partner (CP)</td>
<td>20</td>
<td>22</td>
<td>27</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>User Access Plan</td>
<td>12</td>
<td>17</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Demosing Plan</td>
<td>14</td>
<td>24</td>
<td>21</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>Desired Impact</td>
<td>18</td>
<td>21</td>
<td>23</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>User</td>
<td>10</td>
<td>17</td>
<td>19</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>Root Causes</td>
<td>14</td>
<td>24</td>
<td>25</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td>Value Proposition (VP)</td>
<td>12</td>
<td>20</td>
<td>18</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Existing Solutions</td>
<td>8</td>
<td>15</td>
<td>24</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Implementation Strategy</td>
<td>11</td>
<td>22</td>
<td>22</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>Impact</td>
<td>24</td>
<td>26</td>
<td>26</td>
<td>30</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 4: Expert and novice models of risk prioritization and goal setting

<table>
<thead>
<tr>
<th>Planning Step</th>
<th>Expert Model</th>
<th>Novice Challenges in Risk Prioritization (frequency observed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Prioritization</td>
<td>What experts do: Experts review the risks they identified, prioritize those risks, and select a set of high-priority risks they can address in 1 iteration (1-2 weeks). Often, experts look for economies of scope using a concept called slicing, which involves creating a single plan to address multiple risks across the problem. Further understanding of slicing is crucial, but beyond the scope of this paper. Why it matters: Selecting the greatest risks from across the problem space each iteration allows experts to shift their efforts in response to new knowledge. This minimizes their chances of failing to solve the problem.</td>
<td>Do not attempt to select high-priority risks 10 Spontaneously generate risks (rather than selecting risks identified by analyzing the problem space) 20 Do not prioritize selected risks (among the high-priority risks, which are the most important to address?) 25</td>
</tr>
<tr>
<td>Goal Setting</td>
<td>What experts do: Experts construct a goal that will guide the rest of their planning. This goal takes the form of a conclusively answerable question, or a falsifiable hypothesis or design argument. Why it matters: Having the right goal is crucial because goals inform planning. Useful goals involve resolving questions and hypotheses about important unknowns in the problem space. Experts use such goals to craft plans that will reduce the biggest risks threatening their success.</td>
<td>Goal is not framed as a question, hypothesis, or design argument 23 Goal is too vague to conclusively reach it (e.g., hypothesis is not falsifiable, question is not answerable) 25 Setting a goal that will not reduce the selected risk(s) 15</td>
</tr>
</tbody>
</table>

Externalizing students’ risk analysis using problem space templates

Recall that we designed 2 tools to surface teams’ reasoning about risk analysis: the Design Canvas (DC) and Iteration Plan (IP). These tools are necessary because when an instructor or coach advises a team on their iterative planning, they typically ask a series of questions to surface the team’s reasoning. Likewise, to assess the reasoning behind students’ plans, we need a way to surface that reasoning. In risk analysis, this includes design teams’ knowledge of the problem space, such as their knowledge about existing solutions to the problem (Table 1). It also includes how design teams analyze this knowledge to identify risks in the problem space, select high-priority risks to address, and then construct near-term goals that will reduce those risks.

ICLS 2018 Proceedings 357 © ISLS
Students planned each week by filling out the DC and IP. The final version of the DC works by externalizing what students know about the problem, and what risks they have identified. The DC is a poster-sized template with boxes for each of the key schema components we identified as central to how the expert designers assessed risk in the problem space (Table 1). Students worked in their design teams to fill in the DC by placing sticky notes in each box to represent their knowledge about different aspects of the problem space. Each box in the DC also contained a section for risks, where students placed sticky notes identifying risks in that aspect of the problem space. In a subsection of the risks section, students could rate the risk level as low, medium, or high.

The final version of the IP works by externalizing what students plan to do next, and why they think it is important. The IP is a poster-sized template with boxes for each key component of students’ plans. The IP included boxes for Selected Risks, Goals, Design Method, Metric, Criteria, and Tasks. For assessment, we focused only on Selected Risks and Goals because these determine whether students have begun crafting a plan that links back to major risks in the problem space. In other words, this is the heart of iterative planning. To be effective, students also need to construct coherent plans; they need to choose the right methods, set relevant metrics and criteria, and accurately plan specific tasks. We have evidence that students struggled with these steps, but here we focus only on risk analysis; a discussion of how experts and novices construct coherent plans is outside the scope of this paper.

Assessing students’ risk analysis by using a rubric on problem space templates

Once students have externalized their reasoning using the DC and IP, we can assess it using a rubric that sensitizes assessors to the concepts that experts use in iterative planning, and then directs assessors to check for specific, common novice challenges. The rubric has 3 sections. The first section asks assessors to check the risk in the problem space, and supports this by providing assessors with heuristics for risk detection taken from our expert model (Table 1). This is necessary before assessors can judge how well students have assessed risk in the problem space. The second section asks assessors to check how well students have assessed risk. Assessors check for each of the challenges listed in our novice model of risk detection (Table 3). The third section asks assessors to check how well students have selected risks and set goals. Assessors check for each of the novice challenges listed in our novice model of risk prioritization and goal-setting.

The assessment tools were reliable when applied by different experts with similar disciplinary knowledge. We achieved a Cohen’s kappa = .82 between two researchers testing the final rubric on 5 plans (17% of plans). For each criterion, we coded the plan “yes,” “no,” or “can’t grade.” Inter-rater reliability indicates that the rubric provides sufficient guidance for experts to make similar judgements about quality of students’ plans. Our primary goal was developing expert and novice models and assessment tools. Therefore, we analyzed all available data as we collected it to iteratively improve the expert and novice models and the assessment tools. We then calculated inter-rater reliability by re-analyzing a subset of that data. This is a limitation that we will address by testing the tools on fresh data in future work. Nonetheless, we argue that we can have reasonable confidence that the tools are reliable because 2 researchers coded the 5 plans independently, we had not discussed specific coding decisions when we initially analyzed the data, and this was the first time both of us applied the final version of the rubric to data.

Discussion and conclusion

This paper contributes: (a) a model of how experts in social impact design used risk analysis to inform iterative planning; (b) a model of novice challenges in risk analysis in social impact design; and (c) tools for surfacing and assessing novices’ risk analysis in social impact design.

Our expert and novice models identify risk analysis: an important reasoning process that drives iterative planning, and therefore strategic iteration. Our findings build on previous research which showed that experts use iteration to solve design problems, but did little to define strategic iterative planning, or novice challenges in iterative planning specifically enough to guide assessment (Adams, Turns, & Atman, 2003; Adelson & Soloway, 1985; Atman et al., 2007; Crismond & Adams, 2012; Guindon, 1990). We found that strategic iterative planning requires risk analysis, in which designers identify risks in the problem space, prioritize those risks, and construct goals to reduce them. Likewise, by showing that novices struggle with each step in risk analysis, we provide a plausible explanation for why novices tend to under-utilize iteration (cf. Crismond & Adams, 2012). Also recall that our normative model of iterative planning seems to define practically useful skills (Rees Lewis et al., 2018; see Data Collection and Iterative Analysis section).

We have also demonstrated that it is possible, despite the ambiguity and subjectivity of design problems, to assess risk analysis reliably using tools for externalizing and evaluating students’ reasoning across each step of risk analysis. We have shown that it is possible not only to assess students’ reasoning about their
final solutions (cf. Jonassen, 2010), but also their reasoning about actions taken throughout problem solving—in this case, reasoning about risk analysis. Additionally, while previous work proposed using argumentation rubrics to assess students’ reasoning (Jonassen, 2010), we add that it helps to create templates (the DC and IP) based on expert schemas to externalize the pieces of students’ reasoning you wish to assess. Future work should test the fidelity with which the DC and IP represent and convey students’ thinking to assessors.

Finally, our tools could support assessment of risk analysis in other disciplines (e.g. science or civics). While we designed tools for assessing risk analysis in social impact design problems, risk analysis does not seem specific to social impact design; rather, its function is making highly complex, ill-structured problems more manageable. Thus, it seems like a critical reasoning process for real-world problem solving in any domain.

In this study, we developed tools for assessing students' planning in learning environments for real-world problem solving. Across domains—from science inquiry, to engineering, to civics—learning scientists are committed to designing learning environments for teaching the highly complex, ill-structured problem solving that constitutes professional practice. This includes designing thoughtful assessments that provide instructors with critical insights on where students need help, but the challenges of assessing real-world problem solving are daunting. Our tools overcome these challenges to enable instructors to understand where students are struggling in the critical reasoning process of risk analysis. This may make it easier for instructors to help students develop the knowledge and skills to design solutions to real-world problems.

References

Acknowledgements
We thank Delta Lab, Haoqi Zhang, and Alex Sher for their feedback and support. This work was funded by U.S. National Science Foundation grants IIS-1530883 and IIS-1320693.
A Study of the Design and Enactment of Scientific Modeling Tasks to Support Fourth-Grade Students’ Sense-Making

Meredith Baker Marcum, University of Michigan, merbaker@umich.edu

Abstract: A major goal in education is to provide students with opportunities to engage in “deep learning,” in which knowledge and skills transfer to novel contexts. This study took place in one 4th grade classroom and drew on design-based research methods to investigate how the design and enactment of scientific modeling tasks in a project-based science unit influenced 4th grade students’ sense-making regarding the concept of energy transfer and their epistemological understandings about models and the practice of scientific modeling. Data sources for this study included field notes, video, student artifacts, assessments, and student interviews. Students developed emergent understandings of the concept of energy transfer and the role of models to test, refine, and communicate ideas. This study has implications for the design and incorporation of scientific modeling tasks in curricula, as well as the field’s understanding of the development of upper elementary students’ sense-making regarding the concept of energy transfer.

Keywords: Sense-making, Scientific Modeling, Project-based learning

Research question and theoretical framework

A major goal of education is to support students to transfer the knowledge they construct from the classroom context to future civic, workplace, and family contexts, referred to as “deep learning” (NRC, 2012). The Next Generation Science Standards (NGSS) responded to this call by proposing standards for science education that included not just content knowledge, but also interpersonal practices such as collaboration, as well as other practices identified as characteristic of science that would support students to experience scientific phenomena in authentic and meaningful ways (NGSS Lead States, 2013).

One practice identified by NGSS is scientific modeling. The research question guiding this study was, How do features of the design and enactment of scientific modeling tasks influence 4th grade students’ sense-making of the concept of energy transfer? Sense-making is a cognitive process that could support students in deep learning. Scientific sense-making is a dialogic process that involves both the construction and critique of claims in the pursuit of the construction and refinement of scientific knowledge (Ford, 2012). Engaging in scientific sense-making allows students to participate in talking and reasoning characteristic of scientific practice (Ford, 2012).

Models are critically important in the construction, development, and communication of scientific knowledge (Gilbert, 2004); they serve as simplified representations that allow us to describe, predict, and explain natural phenomena by highlighting specific elements of the phenomenon and relationships among the elements that the modeler considers important in order to explain the behavior of the phenomenon and the associated underlying processes (Gilbert, 2004; Schwarz et al., 2009). This property of models can make them difficult to interpret because they often do not share literal similarities with the phenomenon of interest (Lehrer & Schauble, 2012). Instead, they serve as analogs to the phenomenon representing the underlying mechanisms (Lehrer, Schauble, & Lucas, 2008). Models allow children to represent, test, and revise their thinking in a concrete way, and then interact with that representation to develop their conceptual and epistemological understandings (Penner, 2000). In addition, engaging in the practice of modeling provides opportunities for both individual and collective sense-making about scientific phenomena (Ford, 2012; Manz, 2012).

Historically, the practice of scientific modeling has been considered too complex for young children (Manz, 2012); however, empirical studies have demonstrated that elementary-aged children can be supported to engage in the interpretation of models and the practice of modeling in productive ways (Ero-Tolliver et al., 2013; Lehrer & Schauble, 2012; Lehrer et al., 2008; Louca & Zacharia, 2015; Manz, 2102; Wilkerson-Jerde, Gravel, & Macrander, 2015; Zangori, Vo, Forbes, & Schwarz, 2017). Students’ construction of models can take many forms, including paper-and-pencil, physical objects, computer animations, or computerized simulations. Exposing elementary-aged children to multiple types of models when investigating a phenomenon could support them to develop their “representational repertoire” (Ero-Tolliver et al., 2013, p. 2151), an epistemological understanding that different types of models provide different kinds of information (Schwarz et al., 2009), and make sense of the phenomenon being modeled (Manz, 2012). Building physical objects can provide an “entree” into modeling processes because they retain literal similarities between the model and the target phenomenon.
(Lehrer & Schauble, 2012, p. 174), resulting in lower cognitive requirements. In addition, computer animation supports children to consider the processes at play in phenomena (Wilkerson-Jerde et al., 2015). The use of different forms of scientific models can serve as tools to support students’ sense-making about phenomena and the epistemological aspects of using and constructing scientific models.

As stated previously, sense-making is a social process that requires engagement in the practices of science (Ford, 2012; Manz, 2012). Due to the emphasis on engagement in the practice to support sense-making, I drew on the social constructivist theory of learning to study how engagement in the practice of modeling and the use of models influences students’ sense-making about energy transfer. According to Vygotsky (1978), learning is a social process that is context-specific, mediated by psychological tools, and results in changes in an individual’s psychological development (Palincsar, 1998). Models can serve as a tool that mediates interactions between students and the teacher and externalizes students’ ideas for collective sense-making. In addition, norms regarding how models are used in science have developed over time through the use of models within scientific communities. By engaging students in the practice of scientific modeling, they have the opportunity to engage peripherally in the scientific community by interacting with one of the tools that this community uses.

The use of physical and computer-animated models to support students in making sense of phenomena can also provide opportunities for them to develop epistemological understandings about models and modeling. Given the benefits of engaging students in the practice of modeling to promote deep learning, it is useful to investigate the different features of the instructional context that support productive engagement in the practice of modeling and the use of models to support students’ sense-making.

**Significance**

This study contributes to the field’s understanding regarding how engaging students in the practice of scientific modeling through the use and construction of physical and computer-animated models may support elementary school students’ sense-making about the concept of energy transfer. In addition, this study builds on previous empirical work by studying the development of students’ sense-making regarding the concept of energy transfer across multiple phenomena. This study also demonstrates how theories of learning and development can be applied in an educational context to better understand how digital tools and instructional strategies can support the development of students’ sense-making.

**Research methods**

**Participants and context**

This study took place in a K-5 elementary school, in which 63% of students are White, 24% are African American, and 6% are Hispanic/ Latino; 66% of students qualify for free or reduced lunch; and only 10% of 4th grade students demonstrated proficiency or advanced proficiency on state English Language Arts achievement measures. The participants were Ms. White (a pseudonym) and her 35 fourth-grade students. The racial and socioeconomic makeup of this class mirrored that of the school.

Ms. White enacted a year-long project-based learning (PBL) science curriculum as a part of the Multiple Literacies in Project-Based Learning Project. This project has designed 3rd and 4th grade project-based learning science curriculum that integrates science, math, and language literacy opportunities and aligns with NGSS, and select English Language Arts and Mathematics Common Core State Standards. PBL engages students in meaningful, authentic inquiry and provides opportunities for participation in the practices of science, including modeling, with the goal of supporting students to develop deep conceptual knowledge and epistemological understandings about how scientific knowledge is constructed and critiqued (Krajcik, McNeill, & Reiser, 2007). Two prominent features of PBL are the use of a driving question to frame the unit of inquiry and the creation of student artifacts throughout the unit.

The data for this paper were drawn from one of three fourth-grade units. This unit was framed by the driving question, *Where does the energy to light my home come from?*, and focused on sources of renewable energy and the energy transfers that occur in order to change moving energy of water and wind into electrical energy. The unit was designed to build from students’ experiences with the energy of moving water (that shapes the land) during the previous unit of instruction. Students built physical models of a water wheel and windmill and constructed computer-animated models to explain how each of these devices could be used to harness the moving energy of either water or wind to generate electricity. Students were introduced to the concept of a generator and used hand-crank generators to model the components necessary for the conversion of moving energy to electrical energy on a larger scale. When constructing their physical models of windmills, students used an empty thread spool to model the process of spinning a generator using energy from wind.
The topic of renewable energy was chosen as a way to engage students in investigations, discussions, and experiences related to the concept of energy transfer between objects. In addition to building physical models, students created computerized models using Collabrify Flipbook, a drawing and animation software application, that allowed students to draw multiple frames that could be animated. The animation feature of Collabrify Flipbook supported students to show a process or change over time.

The data sources for this study include video and field notes of whole-class instruction and discussions, students’ scientific models, pre- and post-unit assessments, and interviews with a subset of Ms. White’s students. Six focal students were selected for interviews prior to the start of the unit, at the end of each minicycle, and after the completion of the unit (see Figure 1); they were chosen to represent a range of prior knowledge of concepts related to energy transfer and renewable energies.

The unit pre- and post-assessments included the prompt “Choose one source of renewable energy and draw a model to explain how this source of renewable energy can be used to provide electric power to your home.” During pre- and post-unit interviews, students were asked to explain their models and were asked follow-up questions to further elicit their sense-making about the concept of energy transfer and their epistemological understandings about models and the practice of modeling. During the post-unit interview, students also engaged in a transfer task. The transfer task consisted of a short video of a person using a pool stick to hit a pool ball that strikes another pool ball and were asked the question, “What causes the pool balls to move?” Data from the transfer task were used to capture students’ abilities to transfer their sense-making about energy transfer to a novel context.

Data analysis
To address the research question, this study employed design-based research (DBR) methods to inform revisions to the curriculum. DBR methods seek to design instruction that is theoretically grounded but feasible to enact in the classroom (Brown, 1992; Collins, Joseph, & Bielaczyc, 2004). Therefore, the context within which the instruction occurs is critically important to understanding how learning is actualized in a classroom setting.

Consistent with the design-based research approach, I conducted two minicycles during the unit to study how Ms. White enacted proposed changes to the curriculum and how these changes influenced students’ sense-making, in order to inform future change to the curriculum. Cobb, McClain, & Gravemeijer (2003) define minicycles as short cycles of enactment and reflection within the design experiment, often on a daily basis, that lead to pragmatic changes in order to influence learning during the next instructional opportunity. In contrast, macrocycles encompass the entire design experiment; in this case, the entire enactment of the unit, and analyses of this cycle led to changes in the next iteration of enactment and claims regarding the development of students’ sense-making with scientific models. The analyses of minicycles can document the research team’s learning during the entire macrocycle and contribute to understanding reasons why the macrocycle unfolded in a specific manner (Cobb et al., 2003).

In this study, each minicycle was centered around a different model with which students were engaging. Before and during each minicycle, I worked with the other researchers and Ms. White to discuss sources of struggle and success in students’ sense-making related to the concepts that students were grappling with while engaging with each model, and made adjustments to the activities students were engaged in or provided additional supports that Ms. White could use. In order to study how the design and enactment of scientific modeling tasks influenced students’ sense-making regarding the concept of energy transfer, video and field notes were used to construct a timeline of the unit’s enactment and characterize: (a) the conceptual and epistemological ideas introduced during whole-class instruction, (b) the nature of the supports related to scientific modeling that Ms. White provided to students, and (c) the ways in which the teacher and the students interacted with different models. The close study of students’ development related to the concept of energy transfer and their interactions with different models and the development of their sense-making across the timespan of the unit were informed by the interviews with focal students at different times in the unit and the pre- and post-unit assessments (see Figure 1).
Findings

Minicycle #1

The enactment of this unit took a total of 25 instructional days. During Minicycle #1, students investigated how the energy of moving water could be harnessed to generate electricity by using a physical model and constructing a computer-animated model of a water wheel. For example, the researcher (who served as a co-teacher) made the decision to emphasize comparisons between models and the phenomena they represent. In addition, the researcher and Ms. White decided to ask students to draw models to explain how a water wheel, as opposed to a hydropower plant, could be used to produce electricity, because students had been working with physical models of water wheels for several classes. The objective was to support student sense-making with the water wheel model and introduce the concept of energy transfer, within the water wheel system, to generate electricity.

During whole-class instruction in Minicycle #1, the class looked at pictures and videos of real water wheels and compared them to their water wheel model. This excerpt from a class-wide discussion led by the researcher (MM) illustrates the connections students made between the model and the phenomenon and the struggles they faced regarding their interpretations of models as an analog for the phenomenon. To maintain confidentiality, focal students were assigned pseudonyms and other students were identified using their initials.

Caleb The pipe is like making water, it’s kinda the same because the cup is putting water on our wheel too.
IP It’s kind of like the water wheel, but the water is going onto the ground.
Serenity It is the same shape, but they both have the panels that carry the water.
MM What were the panels on your water wheel? What were they made out of?
Caleb Plastic.
[...]
MM We talked about this during Unit 1, you use these to study things, it starts with an M…
Ashley Model.
MM So your water wheel is a model and what are we using these to study?
Serenity Water wheels
MM But we are learning about something bigger than water wheels.
Madison Energy!

Five of the six focal students included electrical energy as a product of the turning water wheel. Caleb was the only focal student to include and discuss the generator in his model and interview (see Figure 2). He explained,

So it is showing the water wheel and it’s producing electricity to light up the light bulb. The electricity is getting further and getting to the light bulb, the electricity got to the light bulb and lights it up, the rest just shows what everything is…The generator, um, when this moves the generator is getting power from it and it’s turning and making the electricity to light up the light bulb (Caleb Unit 2 Interview #2).

Caleb’s explanation of his water wheel model suggests that the class’s experiences with the physical water wheel model and hand crank generator supported his sense-making and helped him to understand the process of generating electricity. In addition, his explanation of his model suggests that he understood that the turning of the water wheel caused the turning of the generator. This idea could serve as a foundation for energy transfer, but it is unclear how he was thinking about the role of energy in this process because he did not include it in his explanation. Furthermore, he specified that the turning of the generator is what produces the electricity. Caleb’s specificity in the language he used regarding his model was also apparent when he described that his model explained “how the water wheel is turning and making the electricity go to the light bulb.” This explanation of his model suggests that Caleb was developing an emergent understanding of the process of generating electricity from moving water.

Many of the focal students demonstrated that they were considering the conventions of their models and how they could use conventions to improve their models’ communicative power. All focal students used
yellow to denote electricity (as can be seen in the lightbulb in Figure 2). Several students also provided reasons for this choice related to the common convention of using yellow to show electricity. For example, Nick explained that he used yellow, “cause yellow is the general color of electricity, that’s how people see it” (Nick Interview #2). In addition, Esther included arrows to show the direction of the movement of energy. These elements of students’ models suggest that they were considering how to use their models as explanatory tools and the ways in which components within the models could increase their explanatory power.

Figure 2. Screenshot of Caleb’s Water Wheel Model.

The use of Flipbook also provided students the opportunity to use animation. All students created multiple slides for their models, but not all students used animation in purposeful ways to increase the explanatory power of their models. For example, Kiara created three slides, however slides 1 and 2 were exactly the same and slide 3 was a different scene. In contrast, Alyssa used animation to show how the water moved down the water wheel and Esther showed the path of electricity traveling to the house. The use of animation provided the opportunity for students to demonstrate their sense-making with respect to the mechanism through which energy from moving water was transferred to the water wheel and harnessed to generate electricity.

Students’ individual models and interviews also provided insight into their struggles related to their sense-making of how the energy in moving water can be harnessed and used to generate electricity. Students’ responses suggest that the process of energy transfer between moving water and the water wheel was still ambiguous for students. Similarly, students did not understand how the mechanical energy in the turning water wheel was transferred to the turning generator. In addition to not discussing the process of energy transfer, students also did not include representations of energy transfer in their models. This could have been due to the fact that energy is an invisible component and students were struggling to think about the presence of energy throughout the entire system. In Minicycle #2, the researchers made design decisions to address students’ struggles with the process of energy transfer.

Minicycle #2
In Minicycle #2, students investigated how the energy from wind could be harnessed to spin the generator and generate electricity. Students continued to engage with the concept of energy transfer; however, instead of energy transferring from moving water to a spinning wheel, energy was transferring from moving air to a spinning wheel. In the design of the curriculum, these two types of renewable energy sources were chosen due to the overlap in the mechanisms through which energy was transferred and electricity was generated. This minicycle began on Day 18 of Unit 2 and lasted six days of instruction.

During this minicycle, students worked in small groups to plan, build, test, and revise their windmill models. On Day 18, Ms. White introduced students to the representation of the generator that they would be using to test the efficacy of their windmill designs. She explained, “[the generators] won’t actually work, but you have to find a way to have your wind turbine hook up your wind turbine to this generator. You’re also going to be testing your designs. It will be similar to what we did with our water wheels...This spool will be a stand-in for our generator” (Day 18 video). The use of the thread spool was intended to support students’ sense-making about the function and motion of the generator. Ms. White’s introduction of the thread spool introduced the idea that another object could represent the phenomenon being modeled (e.g., a spinning generator). Unlike the water wheel minicycle, Ms. White did not conduct whole class instruction during this minicycle. Instead, she circulated around the classroom and supported students in building, testing, and revising their windmill models.

Focal students demonstrated development in their understandings about energy transfer. All focal students referenced either the process or the effects of energy in the wind being transferred to the windmill. For example, Alyssa described how energy is transferred from the spinning windmill to the spinning generator. These two examples illustrate different points on a continuum of sense-making regarding energy transfer in the windmill system. In addition, only Nick and Esther referenced energy in the wind during their interviews. The
invisible nature of both wind and energy could account for students’ difficulty in understanding that energy was present in the wind and was transferred to the windmill.

After constructing physical models of windmills, students created models in Collabrify Flipbook to explain how their windmill design could light a light bulb. Students’ Flipbook models reflected different levels of abstraction from students’ windmill designs to the phenomenon of harnessing the energy of wind to produce electricity. Some focal students, such as Kiara and Alyssa, created models that reflected their designs and did not include any additional components. Caleb and Nick’s models incorporated some specific components from their windmill designs, but also included components that referenced the phenomenon of lighting the bulb. Caleb’s interview suggests that he was thinking about ways to revise his model to reflect the target phenomenon (i.e., the inclusion of a generator and representations of electricity). Esther’s model and her explanation did not reference her specific design, but was generalized to explain the phenomenon of lighting the bulb (see Figure 3). Despite missing components, such as a generator, Esther’s explanation and the question for her model (How does electricity get to somewhere else?) suggested that she was using her model to explain her understanding of how energy from the wind can be harnessed to produce electricity, as opposed to explaining her windmill design. These examples illustrate different levels of students’ understanding regarding how models can be informed by data and experience to explain a larger, more abstract phenomenon.

The analysis of Minicycle #2 revealed the ways in which they could draw on their own investigations and experiences to inform the construction of models, which represented a larger or more abstract phenomenon. All focal students, except Kiara and Samuel (who did not have a model), included wind as a component in their models. While people can feel wind, and observe the effects of its movement, it is impossible to see wind itself, unlike being able to observe the movement of water and the moment of impact when water collides with another object, such as a water wheel. The invisibility of wind was another aspect that made the phenomenon of the transfer of energy from wind more challenging for students to make sense of in contrast to the transfer of energy from water.

The concept of a generator and its role in the process of energy transfer and electricity production remained a difficult concept for focal students to make sense of, despite the prominent role that the generator played in the physical windmill models that students constructed and tested. The number of students who incorporated the generator into their model remained consistent from the water wheel models. Caleb, Nick, and Alyssa incorporated the generator into their models or their discussions about what they would add to their models, while other focal students continued to reference electricity as emerging from the windmill. The other focal students indicated that electricity emerged directly from the water wheel. These continued challenges underscore the difficulty of this concept.

Students’ contributions during whole class discussions suggested that some students were thinking about energy, but, similar to the conversations regarding the water wheel, there were no sustained discussions involving energy transfer. Because of the materials students were using, there were fewer similarities to real windmills than when students engaged with the water wheel model, resulting in higher cognitive requirements for students to make sense of this model (Lehrer & Schauble, 2012). This was also their first experience with a syntactic model, or a model that maintains functional similarities and not physical similarities with the phenomenon of lighting the bulb (Lehrer & Schauble, 2012).

**Macrocycle analysis**

In order to determine students’ endpoints in the unit and the development in their sense-making regarding energy transfer, I analyzed focal students’ post-unit interviews and assessments and the two minicycles to make claims regarding how the curriculum may have supported their sense-making. Five of six focal students identified that energy was present in the renewable source of energy that they chose to model. For example, Alyssa, Nick, and Samuel drew models of windmills and identified that the wind provided energy for the spinning windmill, while Esther and Caleb drew solar panels and identified the sun as providing energy to the solar panels. Kiara was the only focal student who did not include a representation of the energy source or provide insight into where the energy came from in her post-unit model of a windmill. These five students also
included representations of wind or solar energy in their models, suggesting that they had developed ways of including invisible components and felt they were important to include in their models. The development of this understanding can be observed across the two minicycles and may have been supported through Ms. White’s prompts to students regarding using their models to answer questions, as well as the sharing of students’ models with the class.

Students also demonstrated a range of understandings regarding energy transfer in the transfer task during the post-unit interview. For example, Esther was able to describe, in detail, the path of energy to explain why the pool balls move. She explained, “From the person hitting it from the stick. And the energy from the person goes to the stick and when the stick hits the ball the energy from the ball goes to the ball that it hits” (Esther Unit 2 Interview #5). However, Caleb was not able to use apply the word “energy” in the transfer task and, instead, drew comparisons between what he observed in the video and a previous investigation, students had conducted using rubber bands to launch steel balls and observed the effects of their collisions with a container as an introduction to energy transfer. Caleb’s comparison suggests that he saw similarities between the two contexts, but was not able to apply the concepts he used to analyze his data from the investigation to the novel context of the pool ball task (Wagner, 2010).

Students continued to experience challenges in understanding the role and function of the generator. Despite the emphasis placed on the generator during Minicycle #2, not all focal students included generators in their Flipbook models at the end of Minicycle #2. Three students included generators and two students did not (Samuel did not construct a windmill model). In addition, on the post-unit assessments, no focal student included a generator in their model. This could have been due to the prompt, but also suggests that students still did not consider the generator to be a necessary component in an explanation of generating electricity.

Conclusions and implications
The aim of this study was to investigate how the features of the design and enactment of scientific modeling tasks influence 4th grade students’ (a) sense-making of the concept of energy transfer and (b) epistemological ideas related to the use and construction of scientific models. Using DBR methods, analyses of the minicycles within the unit and macrocycle of the entire unit revealed how the design of the use of physical models and the construction of computer-animated models influenced and provided insight into students’ sense-making regarding energy transfer.

Students demonstrated increased understanding regarding the use of a generator in the process of generating electricity, but several students did not include generators in either their drawn water wheel or windmill models. The use of the thread spool to represent the generator in the physical windmill model was designed to support students to understand the centrality of the generator to the process of generating electricity and to understand the energy transfer necessary in order to make the generator spin. While some students did include generators in their models or discussed their models during their generators, this was not true for all students. The thread spool did not share any resemblance with a generator that students had observed and instead served as an analog to the generator (Lehrer et al., 2008). The abstractness of this representation may have limited students’ abilities to make connections between the process of spinning the generator and spinning the thread spool. Providing additional supports in the curriculum that teachers could draw on to help students identify similarities between the processes of the thread spool and the generator, as opposed to literal similarities, could support students to incorporate the role of the generator in their models and explanations.

A comparison of the two minicycles revealed that more whole class discussions regarding the ideas related to energy transfer occurred during Minicycle #1 as compared to Minicycle #2. One reason for this was that the length of Minicycle #1 was longer than Minicycle #2, and therefore provided more opportunities for whole class discussions. Despite a greater number of whole class discussions in Minicycle #1 than in Minicycle #2, there were no sustained conversations about energy transfer during either minicycle. The lack of opportunities for the class to introduce and build on each other’s ideas related to energy transfer may have limited the development of students’ sense-making regarding this concept. These comparisons suggest that more prompts for whole class discussions should be integrated into the curriculum to support teachers to initiate conversation regarding the concept of energy transfer as it relates to the modeling tasks in which students are engaged. In addition, providing opportunities for teachers to develop their content knowledge regarding the concept of energy transfer during professional development could allow them to better support students’ sense-making about energy transfer through class discussion and the use of models.

The use of Collabrify Flipbook supported students to demonstrate their sense-making regarding the harnessing of energy from water or wind to generate electricity. In both models, multiple focal students used animation in their models to show the process of water or wind moving and the resulting movement of either the water wheel or windmill. The use of Flipbook also provided opportunities for students to create their own
representations of the process, including conventional ways to show electricity that others would understand and the incorporation of invisible components (e.g., wind). These features supported students’ sense-making regarding both conceptual ideas related to energy transfer and epistemological ideas related to the practice of scientific modeling.

References

Acknowledgements
The research and development reported in this paper is being supported by a generous grant awarded to Joseph Krajcik, Annemarie Palincsar, and Emily Miller from the George Lucas Educational Fund entitled, Multiple Literacies in Project-Based Learning. Any opinions, findings, and recommendations expressed in this paper are those of the author. Special thanks to Joseph Krajcik who led the design of the unit of study central to this paper.
Examining Pre-service Teacher Knowledge Trajectories of Computational Thinking Through a Redesigned Educational Technology Course

Hui Yang, Chrystalla Mouza, Yi-Cheng Pan,
Email: huiy@udel.edu, cmouza@udel.edu, ethanpan@udel.edu
University of Delaware

Abstract: Computation Thinking (CT) is a fundamental skill of analytical thinking. To successfully infuse CT into K-8 settings, we must equip pre-service teachers with CT knowledge and skills that can be applied in their curricular context. This work presents trajectories of pre-service teacher knowledge development of CT concepts and computing tools within the context of disciplinary content and pedagogical knowledge through their participation in an educational technology course. Data were collected from a random sample of 42 pre-service teachers. Data sources included participants' lesson plans (N=126) collected through three different time points. These materials were analyzed quantitatively and qualitatively. Preliminary results indicated that pre-service teachers gradually improved their understanding throughout their participation in the course. Findings also revealed difficulties in conceptualizing and integrating CT in conjunction with content and pedagogy, particularly during authentic vis-à-vis hypothetical lesson designs.

Introduction
In recent years, researchers and policy makers have established a compelling rationale for introducing computing in K-8 contexts as a means of advancing student development of Computational Thinking (CT). Broadly speaking, CT is a problem-solving methodology that can be implemented with a computer and can be automated, transferred, and applied across subjects (Barr & Stephenson, 2011). Wing (2006) suggests that CT is “a universally applicable attitude and skillset for everyone, not just computer scientists …” (p. 33). Accordingly, the International Society for Technology in Education (ISTE) in collaboration with the Computer Science Teacher Association (CSTA) identified CT concepts for K-12, which include (a) problem decomposition as “breaking down complex problems into more manageable parts”; (b) algorithmic thinking as “using a precise sequence of steps or instructions to solve problems”; and (c) simulation as “representing a process” (CSTA & ISTE, 2011). Indeed, the newly released National Educational Technology Standards for Students emphasized the need for advancing our younger generations to develop CT skills required to navigate in the digital world (ISTE, 2016). Promoting the development of CT knowledge and skills in K-8 settings, however, is challenging because teacher preparedness is a major barrier (Code.org, 2017). Thus, a critical step for successfully infusing CT into K-8 classrooms is to help pre-service teachers build an understanding of CT and its connection to their curricular context (Yadav, Hong, & Stephenson, 2016; Yadav, Stephenson, & Hong, 2017).

Since its inception, the framework of technological pedagogical content knowledge (TPACK) has provided a unifying lens for researchers working to explicate teacher knowledge for effective use of technology tools and practices across the curriculum (Mishra & Koehler, 2006). Thus, TPACK provides a useful framework for studying teacher knowledge in relation to CT, because computational tools are central to CT (Angeli, Voogt, Fluck, Webb, Cox, Malyn-Smith, & Zagami, 2016). Research that examines TPACK in relation to CT, however, is only now beginning to emerge (Mouza, Yang, Pan, Ozden, & Pollock, 2017). One way to advance pre-service teacher preparation for integrating CT into K-8 education is through stand-alone educational technology courses required in most teacher education programs around the U.S. (Yadav, Gretter, Good, & McLean, 2017). In this work we present the design of an educational technology course which introduces computing tools, vocabulary, and practices specific to incorporating CT within the context of content and pedagogical knowledge in K-8 settings. Relatively, we explore the following research question: What learning trajectories are exhibited by pre-service teachers as they learn and apply CT-related concepts and computing tools within the context of disciplinary content and pedagogical knowledge?

Theoretical framework
This work is situated in the theoretical framework of TPACK (Mishra & Koehler, 2006). Building upon Shulman’s (1987) scholarship of teacher knowledge, TPACK centers on the nuanced interactions among three bodies of knowledge (Figure 1): content knowledge (CK), technology knowledge (TK), and pedagogical
knowledge (PK). CK refers to knowledge of subject matter. TK refers to knowledge of various technologies and appropriate vocabulary (e.g., terminology). PK refers to knowledge of methods and processes for teaching. These domains combine to form three additional constructs. Pedagogical content knowledge (PCK) refers to knowledge of representing content to make it comprehensible to others. Technological content knowledge (TCK) refers to knowledge of how technology can create new content representations. Technological pedagogical knowledge (TPK) refers to knowledge of how various technologies can be used in teaching. When technology, content and pedagogy blend together, the result is TPACK—a synthesized form of knowledge that supports effective use of technology within specific subject domains.

In prior work (Mouza et al., 2017), we explicated the construct of TPACK in relation to CT, focusing on what all teachers need to know and be able to do in order to use CT as a means for exploring disciplinary content (e.g., math, science, literacy, etc.). Towards this end we advanced the term of TPACK-CT, which focuses on pre-service teachers’ ability to understand how CT-related concepts, computing tools, and practices (TK) can be combined with disciplinary content (CK) and pedagogical strategies (PK) to promote meaningful student outcomes in specific contexts (see Figure 1). The construct of TPACK is used in two ways in this work: as a framework guiding the design of the educational technology course, and as an analytic lens for examining pre-service teacher outcomes as illustrated in course products. We focus exclusively on the construct of TPACK-CT, because our goal is to move beyond the individual knowledge components of technology, content and pedagogy to illustrate a synthesized form of knowledge that supports effective use of CT-related concepts and tools within specific subject domains.

**Figure 1.** TPACK-CT framework (based on Mishra & Koehler, 2006).

**Methods**

**Context**

This study was conducted in the context of a four-year undergraduate teacher education program in the United States. Graduates of the program are eligible for both elementary (K-5) and middle school (6-8) teacher certification. The program curriculum is divided into three areas: (a) the general studies courses which help develop subject matter knowledge; (b) the professional studies courses (e.g., methods) which prepare pre-service teachers for their future classroom; and (c) the concentration courses which help develop expertise in a middle school content area. Additionally, the program curriculum is designed to provide pre-service teachers with a range of field experiences in a variety of classroom settings. These experiences culminate with student teaching.

**Course description**

Integrating Technology in Education is a 15-week course required for all pre-service teachers during their junior or senior year. Typically, the course introduces participants to technologies available for use in classroom content areas, pedagogical considerations with these technologies, and teaching and learning practices that combine the use of technologies with content and pedagogy. The specific tools used in the course change frequently to keep pace with rapid advances in technology, but typically include tools that support communication, content representation, collaboration and production. Concurrent with the course, pre-service teachers complete methods courses and accompanied field experience for 3 full weeks within a classroom setting. The field experience allows the opportunity to engage in the design and application of authentic
classroom materials that embed technologies in the context of content area instruction. For the purpose of this work, we redesigned this course to support the development of pre-service teachers’ use of CT-related concepts and computing tools (TK-CT), as well as to promote their CT practices within specific disciplinary and pedagogy (TPACK-CT). Table 1 provides an overview of the course design. The course is offered every semester but for the purpose of this work we focus on all sections offered during three consecutive semesters (Spring 2016, Fall 2016 and Spring 2017). All sections were taught by the same instructor (second author) and utilized the same syllabus and course activities.

Table 1: Description of the CT-infused educational technology course

<table>
<thead>
<tr>
<th>Technology</th>
<th>Activity</th>
<th>CT Supported Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive Whiteboards</td>
<td>Identify two interactive whiteboard resources that support key CT skills: Modeling (e.g., a resource that can be used to represent a phenomenon such as prey and predator relationship); Sequencing (a resource that could be used to sequence events); Data (e.g., a resource that could be used to represent data such as a graph); and Sorting (e.g., a resource that could be used to organize information)</td>
<td>Modeling, Abstraction, Algorithmic, Thinking, Data &amp; Sorting</td>
</tr>
<tr>
<td>Programming Hour of Code</td>
<td>CS unplugged (activity done without computers to introduce algorithms) Hour of Code: Completion of a Grades 2-8 activity Introduction to Scratch Programming: Scratch is an object oriented programming language Review lessons that support the use of Scratch: ScratchED Design of a learning activity in a content area that involves Scratch Programming/Reflection</td>
<td>Algorithmic, Thinking Problem, Decomposition, Modeling, Abstraction</td>
</tr>
<tr>
<td>Scratch</td>
<td>Concept Mapping Tools: Design of a learning activity that uses concept mapping in a content area to support student development of CT skills, such as decompose a mathematical problem, model abstraction (e.g., life cycle of a butterfly), sequence events in a story or plan essay execution.</td>
<td>Problem, Decomposition, Algorithmic, Thinking, Modeling, Abstraction</td>
</tr>
<tr>
<td>Data</td>
<td>Data: Introduction to Internet Research including use of keywords, boolean logic and operators and evaluation of online content</td>
<td>Problem, Decomposition, Modeling, Abstraction</td>
</tr>
<tr>
<td>Collaboration tools</td>
<td>Collaboration tools: Select and read an article on multiple approaches to developing CT: board games, robotics, programming. Use of a collaboration tool to present the reading to classmates.</td>
<td>Problem, Decomposition, Algorithmic, Thinking, Modeling, Abstraction</td>
</tr>
</tbody>
</table>

Participants
A total of 135 pre-service teachers enrolled in six sections of the course between Spring 2016-Spring 2017. For the purpose of this study, six to eight pre-service teachers were randomly selected from each section for a total of 42 participants (N=42), which represents approximately 30% of all participants.

Data collection
To examine pre-service teachers’ growing trajectories as they represented and applied their TK-CT with content and pedagogy in the context of classroom teaching (TPACK-CT), we collected course materials at three different time points. For time point 1 (T1), we collected pre-service teachers’ lesson plans incorporating a programming tool – Scratch (see Table 1). In this activity, pre-service teachers developed lesson plans that incorporated Scratch programming within a curricular content area. To accomplish this goal, they first examined user-created projects available on the Scratch community that fit their curricular goals. Subsequently, they examined a variety of lesson plans through the ScratchEd community, an online publicly available forum where educators exchange resources, pose questions and share experiences to broaden the integration of Scratch into
core curricular contexts. Finally, they identified a learning goal within a content area of their choice (e.g., science, mathematics, social studies, English) and developed a lesson plan that integrates programming with curriculum content. To scaffold lesson development, pre-service teachers were provided with a series of prompts aligned with the TPACK framework (see Harris, Grandgenett, & Hofer, 2010), including (a) planning (e.g., Consider the pedagogical decisions you’ll need to make to develop this lesson idea. How will you introduce Scratch to your students? What activity types will students engage with to learn the concept? In what ways will you assess student learning?); and (b) reflection (e.g., How will programming help your students achieve the identified learning goal? How will the lesson support the development of students’ CT skills? What was your experience with Scratch?).

For time point 2 (T2), we collected pre-service teachers’ lesson plans incorporating a concept mapping tool. Initially, all pre-service teachers were introduced to various types of concept mapping tools. Much like learning to use programming tools, pre-service teachers were subsequently asked to identify a learning goal within a content area of their choice (e.g., science, mathematics, social studies, English) and develop a lesson plan that integrates concept mapping in a core curricular area. To scaffold lesson development, a series of planning and reflection prompts was again provided that paralleled those of the programming activity.

For time point 3 (T3), we collected pre-service teachers’ final course products completed through a case development project. The case development project progressed incrementally through stages that allowed participants to implement and reflect on their own lessons that supported the development of CT knowledge and skills among school students. Participants were allowed to implement one of the lessons designed throughout their participation in the Technology Integration course (e.g., programming, concept mapping) or design a new lesson that integrated technology with content and pedagogy aimed at supporting students’ CT knowledge and skills. The culminating component of the project was a reflective case report of approximately 1,000 words written in response to several prompts. Specifically, each case report was divided into two sections: (a) case narrative (e.g., How did you introduce the lesson to students? What happened during the actual implementation of your lesson?), and (b) case reflection (e.g., How did the lesson support the development of students’ CT skills? What are two things you will remember about this lesson for future planning?).

It is important to note that the programming and concept mapping activities asked pre-service teachers to develop hypothetical lesson plans, which helped them envision the infusion of CT into their future classroom. In contrast, the final case development activity asked participants to implement and reflect upon an authentic lesson designed and enacted in their field experience. All assignments followed a consistent format; they first engaged pre-service teachers in developing knowledge and skills related to CT concepts and computing (TK-CT) and subsequently asked them to apply such knowledge in the context of disciplinary content and pedagogy (TPACK-CT).

Data analysis
A total of one hundred and twenty-six lesson plans (N=126) were scored using a modified version of the Technology Integration Assessment Rubric, a valid and reliable instrument aligned with the TPACK framework, that can be used to evaluate pre-service teachers’ lesson plans (Harris, Grandgenett, & Hofer, 2010). The rubric identifies four evaluation criteria which include: (a) curriculum goals and technologies (e.g., computing tools and practices that support the development of CT knowledge and skills); (b) instructional strategies and technology: using computing tools to support teaching and learning that fosters students’ CT knowledge and skills; (c) technology selection(s): compatibility with curriculum goals and instructional strategies; and (d) fit: alignment of content, pedagogy, and computing tools to foster CT knowledge and skills. Each of the four criteria can receive a numerical score from 1 to 4. A score of 1 indicates failure in satisfying the criterion, while a score of 4 indicates full success in satisfying the criterion. Each lesson plan was scored by two researchers. The initial inter-rater reliability was calculated at 87%. All discrepancies were discussed until a 100% agreement was reached.

A one-way repeated measures analysis of variance (ANOVA) was conducted to analyze and identify all 42 participants’ TPACK-CT growing trajectory based on their overall scores at three different time points (T1, T2 and T3). Subsequently, participants’ T1 overall scores were categorized in three groups: High Start, Low Start and Medium Start groups. The High Start group (N=5) included participants whose lesson plans scored 3.0 or above (top quarter). The Low Start group (N=24) included participants with lesson plans that scored below 2.0 (bottom half). The Medium Start group included participants (N=13) whose lesson plans received scores between 2.1 to 2.9. Based on initial scores, data were subsequently analyzed using a two-way repeated measures analysis of variance (ANOVA) to present growth trajectories among the three groups, followed by a one-way ANOVA performed at each time point to determine the mean differences among groups. We subsequently selected three participants, who were representative of each group, and conducted a qualitative
analysis of their data entries over time in order to better illustrate how TPACK-CT was represented in these growing trajectories.

**Findings**

**Pre-service teachers’ development of TPACK-CT**

A one-way repeated measures ANOVA with a Greenhouse-Geisser correction determined that means of participants’ overall TPACK-CT scores significantly differed between time points \( F(1.96, 80.52) = 26.97, P < 0.0005 \). Post hoc tests using the Bonferroni correction revealed that participants’ overall TPACK-CT scores significantly increased in the concept mapping lessons (T2) and final case narratives (T3) compared with the Scratch programming design (T1). However, there was a slight reduction, 0.39 point, in participants’ TPACK-CT scores from their concept mapping lessons to the final case narrative, which was statistically significant \( p = .03 \). Figure 2 provides an overall growing trajectory among all participants.

![Figure 2. Plot of marginal means across time (N=42).](image)

Results from two-way repeated measures ANOVA revealed that the group-by-time interaction was also significant (Wilks’ \( \lambda = .42 \), \( F(2, 4) = 10.18, p = .000 \)). Tukey post hoc tests from one-way ANOVA revealed that a statistically significant difference existed among “High Start”, “Medium Start” and “Low Start” at T1 \( (F(2, 39) = 168.16, p = .000) \); however, no significant differences existed among these groups in the concept mapping assignment (T2) and final case narrative (T3) even though they started with significant differences in the programming lesson plan (T1). Table 2 presents the means and standard deviations for the three groups on the dependent variable separated by time period. Figure 3 provides a visual representation of the three growing trajectory patterns.

**Table 2: Means and standard deviations by groups and times**

<table>
<thead>
<tr>
<th>Time</th>
<th>Groups</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>High Start</td>
<td>3.25</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Medium Start</td>
<td>2.46</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Low Start</td>
<td>1.43</td>
<td>0.14</td>
</tr>
<tr>
<td>T2</td>
<td>High Start</td>
<td>3.20</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Medium Start</td>
<td>2.86</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Low Start</td>
<td>3.13</td>
<td>0.69</td>
</tr>
<tr>
<td>T3</td>
<td>High Start</td>
<td>2.83</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Medium Start</td>
<td>2.77</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Low Start</td>
<td>2.57</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Overall, quantitative data indicated a growing understanding of TPACK-CT in pre-service teachers’ lesson plans over time. Specifically, participants who started with a high score on the programming lesson plan maintained high TPACK-CT development. Further, participants who had limited to moderate TPACK-CT understanding in the programming lesson plan gradually improved in their development of TPACK-CT as evidenced in the concept mapping and final case narrative. These findings indicate that participation in the Technology Integration course positively influenced pre-service teachers’ TPACK-CT. However, as shown in our analysis, participants exhibited some difficulty in articulating TPACK-CT in authentic settings (i.e., case narrative) compared to hypothetical lesson design (i.e., concept mapping).

Pre-service teachers’ TPACK-CT representation

To more clearly illustrate pre-service teachers’ growing trajectories, we present a more detailed analysis from three participants who represent the growth exhibited within each group: Jim (High Start group), Casie (Low Start group) and Ella (Medium Start group). Figure 4 presents the TPACK-CT growing trajectories of these three participants.

**Jim: High Start group**

Jim’s TPACK-CT trajectory started with a high numerical value of 3.3 in his programming assignment where he provided a sound explanation around the ways in which his lesson plan can support the development of student CT knowledge and skills. Jim’s programming lesson focused on the integration of Scratch in Language Arts. Specifically, Jim’s plan built on a lesson observed in his field placement where second grade students read about different forms of transportation. To extend the lesson, Jim planned to have students generate a story that uses two different forms of transportation (one modern form and one past form) and subsequently engage in peer reviews and iterative revisions. As a final step, the plan asked students to create an interactive digital story using Scratch. To achieve this goal, Jim proposed to cooperate with the school’s technology teacher to introduce students to block-based programming. Jim’s reason for incorporating Scratch (TK) into his Language Arts lesson was to help students make connections between the writing process and the “sequencing mindset that is needed to program things” (CT). Notably, Jim did not have prior programming experience when he enrolled in the course; however, he demonstrated positive attitudes towards Scratch with accurate understanding around the role of the tool in supporting the integration of CT with content and pedagogy (TPACK-CT).

In his second lesson plan, Jim focused on the use of a concept-mapping tool as a means of helping students visualize and represent two-step addition problems in mathematics. Similar to his programming lesson, Jim proposed to build up student knowledge of using concept mapping tools (TK) while simultaneously helping
them develop conceptual understanding about addition and place values (i.e., content). Subsequently, he planned to provide students a math word problem and ask them to use the concept mapping tool to represent the solution (CT). Jim pointed out that using the concept mapping tool could benefit students’ math conceptual understanding while supporting CT skills such as problem decomposition (TPACK-CT).

Jim was able to provide evidence of TPACK-CT in his programing and concept mapping lessons but not in his final case narrative where he implemented a stand-alone technology lesson with no connection to curricular content. Specifically, Jim’s final case focused on the Hour of Code (see Table 1) as a way of helping students develop CT skills such as sequencing and algorithmic thinking.

**Casie: Low Start group**

Casie’s programming lesson received a score of 1.25, which placed her in the Low Start group. In this lesson, Casie utilized Scratch games into a math lesson with the objective of helping students use place value understanding to whole number to the nearest 10 or 100. The use of programming, however, did not support the lesson’s objectives. Casie simply assumed that playing Scratch games would deliver the math content. After playing the games, she planned to assess students using a worksheet. In this lesson, Casie demonstrated a misunderstanding of CT by asserting that it was the skill that “students would need to understand (when) they need to click on the right answer to complete the lesson”. Reflecting on her experience, Casie noted that she found Scratch difficult and confusing.

In her second plan, Casie demonstrated a developing understanding of TPACK-CT. In this lesson, Casie utilized a concept mapping tool to help students understand a story. Her plan focused on reading the story of the *Three Little Pigs* to her students and subsequently completing a concept map focusing on the “who, what, when and how” of the story. In this lesson, concept mapping can help facilitate students’ CT in the areas of problem decomposition (breaking up the story into smaller chunks) to facilitate understanding of the story. Although Casie provided a connection between the use of concept mapping within a content area, her pedagogical decisions on how to guide first graders use of concept mapping tools was relatively weak.

In her final assignment, Casie used an interactive whiteboard application to deliver a math lesson on fractions. Specifically, she presented 7 brownies and had students come up to the board to equally sort the brownies into four portions. This lesson could support students’ CT knowledge and skills through problem decomposition and simulation with the use of technology. Casie, however, did not recognize these connections and rather indicated that the lesson supported students’ CT because it allowed them to “be hands on with technology”.

**Ella: Medium Start group**

Ella represented the group of pre-service teachers who scored between 2.2 to 2.9 in the programming lesson and steadily advanced throughout the duration of the course. Ella’s programming lesson was designed to extent a math lesson observed in her field placement that focused on helping first graders practice addition and subtraction within 20. The plan engaged students with a series of progressively more difficult games in Scratch where they could practice addition and subtraction followed by completing an exit ticket aimed at checking understanding of relevant concepts. Finally, Ella’s plan involved students creating their own addition and subtraction game in Scratch. Ella’s lesson demonstrated a moderate understanding of TPACK-CT; although she incorporated programming into the lesson, she provided limited explanation of the technological and pedagogical decisions that would help support young students’ math learning through programming. Unlike Casie, however, Ella reported a positive learning experience with Scratch programming.

In her concept mapping lesson, Ella aimed to help students find factor pairs for a whole number in the range from 1 to 100. To achieve this goal, she used a concept mapping tool to help students visually represent and organize math facts and solutions. Besides having whole class instruction, Ella also had students work on creating individual visual representations of pair factors. Ella’s concept mapping lesson demonstrated a developing TPACK-CT since she was able to frame the use of a computing tool with content and pedagogical context. Besides helping students achieve the math learning goals, Ella’s concept mapping lesson supported students’ CT development in problem decomposition, simulation and algorithmic thinking.

Ella’s final product, utilized various forms of technology in conjunction with a social studies concept related to the *Three Branches of Government*. In this lesson, however, the connection between technology (e.g., video and projector), content and pedagogy was relatively weak. The video contained probing questions that helped students break down the meaning of the text (e.g., problem decomposition), but Ella failed to conceptualize her content and pedagogical decisions associated with the use of the technology and how those may support CT.
Conclusion and implications

For pre-service teachers to successfully integrate CT in school curricula, they must develop a sound understanding of CT concepts, computing tools and practices. In this work we examined pre-service teachers’ growing trajectories as they designed CT-infused lesson plans through their participation in an educational technology course offered in conjunction with field experience in authentic classrooms. Examination of pre-service teachers lesson designs indicated a growing understanding of TPACK-CT over time, despite differential starting points. Problem decomposition and algorithmic thinking were the most frequently cited CT constructs in participants’ lesson designs. Overall, participants scored higher in the concept mapping lesson design; they illustrated a good understanding of TPACK-CT by using concept mapping tools to support both curricular objectives and the development of CT (e.g., problem decomposition, simulation, etc.). However, it is worth noting that this assignment was hypothetical – participants did not have to implement it.

In contrast, participants demonstrated a dip in the representation of TPACK-CT in practice. In this assignment, pre-service teachers encountered difficulties in either selecting appropriate computing tools or infusing CT concepts into the context of disciplinary content and pedagogy. Moreover, participants frequently failed to recognize other crucial CT concepts as they delivered their lessons including data analysis and automation. This finding is not surprising given participants’ lack of classroom teaching experience. As a form of knowledge rooted in classroom practice TPACK-CT is influenced by pre-service teachers’ understanding of pedagogy in relation to both content and computing tools (Mouza et al., 2017). For pre-service teachers to successfully build and TPACK-CT and apply it in practice, we must provide extensive opportunities to develop both their theoretical and practical knowledge (Gomez, Sherin, Griesdorn, & Finn, 2008).

References


Interconnecting Knowledge, Experience, and Self in Humanistic Knowledge Building Communities

Yotam Hod, University of Haifa, yotamhod24@gmail.com
Dani Ben-Zvi, University of Haifa, dbenzvi@univ.haifa.ac.il
Michal Dvir, University of Haifa, dvirmich@gmail.com

Abstract: This study attends the learning sciences to the work of Carl Rogers and person-centered education. Building on claims that knowledge building communities are idea-centered, as well as recent research in this area that has looked at learning holistically, we examine the notion of a ‘humanistic knowledge building community’ as an integration of idea-centered classroom knowledge building communities with Rogerian person-centeredness. We investigated an innovative course for graduate students in an educational technologies program that both inspired and informed this conceptualization. The results of our research led to an operational KES framework that reveals different patterns of interconnections between students’ evolving knowledge, experience, and self. In this study we report on these through the lens of a single student.

Introduction and background
The learning community approach has been one of the most significant contemporary developments in educational instruction addressing the vital need for meaningful interpersonal learning in the digital age (Adams, Becker, Freeman, Giesinger, Cummins, & Yuhnke, 2016). Over the past several decades, considerable knowledge has accumulated as to the ways that these classroom learning communities should be designed (Hod, Ben-Zvi, & Bielaczyc, 2016). Knowledge building communities, a paramount model of classroom learning communities, have been idea-centered, with a focus on having participants learn as they take responsibility over advancing their collective knowledge (Bielaczyc, Kapur, & Collins, 2013). In recent years, research has begun to look at the interplay between students’ knowledge, the activities they engage in, and their identities (Herrenkohl & Mertl, 2010). With an eye on contributing to research on knowledge building communities in this direction, in this paper we draw on data from a unique classroom knowledge building community that introduces person-centered activities—rooted in the work of Carl Rogers—to its design. The purpose of this paper is manifold: First, to draw out the differences between idea- and person-centeredness based on their unique academic lineages; second, to show how their integration is without precedent even though several lines of research address different aspects of it; and third, to elucidate how their integration leads to a unique phenomenon we call a ‘humanistic knowledge building community’, which we subsequently instantiate in a case study of a learner.

Idea-centeredness
Knowledge building communities have been at the forefront of the effort to rethink education (Bielaczyc et al., 2013). While in many ways similar to other learning community models, knowledge building communities differ in their emphasis on ‘knowledge work’ over ‘learning’. Stated differently, knowledge building communities foster learning in its participants, but as a by-product of engaging in the enterprise of progressively advancing knowledge. Idea-centeredness refers to a commitment of advancing the collective knowledge of an organization. The implications of being idea-centered can be seen in various aspects of the classroom, such as the types of activities that are designed and the technologies used to support them. Learning communities may involve sharing and whole-group discussions so that students can learn from each other. In knowledge building communities, participants must take collective cognitive responsibility by being aware of others’ contributions, making complementary contributions, and engaging in varied roles necessary to achieve the collective knowledge goals (Zhang, Scardamalia, Reeve, & Messina, 2009). In terms of technology, learning communities offer mechanisms for students to exchange and build on each other’s ideas. In knowledge building communities, technologies such as Knowledge Forum focus students to relate their contributions to those that the community already developed. Not only do certain features point out what ideas they have read or not read, when participants post their ideas they mark the relationship of their ideas (e.g., build-on) to the growing database of knowledge.

This view of idea-centeredness does not mean that knowledge building communities are the only classroom learning communities that are exclusively idea-centered. In a review of several socioculturally minded classroom learning communities, Bielaczyc, et al. (2013) explain that “putting students’ ideas at the center of the community work communicates to students that their ideas matter to others and that they have a position of responsibility in contributing to the community’s advancement” (p. 4). Learning communities are idea-centered...
in so far that the knowledge is public and advancing it is a collective responsibility of its participants. By being explicitly idea-centered and as an innovator of this approach, knowledge building communities are the paragon model of idea-centeredness.

**Person-centeredness**

Central in Carl Rogers’ perspective of person-centeredness is that human life has an inherent motivation to expand and develop. So inherent was this belief that therapists and educators didn’t need to inspire self-fulfillment in patients or students, but rather served the purpose of removing the obstacles that blocked personal growth (Rogers, 1969). Free from interpersonal, societal, and cultural restrictions, people could fully-function or actualize themselves in an ongoing process of self-discovery. So far reaching were the implications of these ideas, they spread into nearly every form of modern organization including the scientific establishment, which increasingly accepted post-positivistic perspectives and methodologies (Rogers, 1970).

Although individual therapy and counseling was the large focus of the early part of Rogers’s career, he was deeply interested in intensive group experiences for a large part of his later career. Influenced by Gestalt Psychology and Kurt Lewin, who opened the Research Center for Group Dynamics at MIT in the 1940s, Rogers championed the encounter group (Rogers, 1970), what he considered to be “perhaps the most significant social invention of this [20th] century” (Rogers, 1968, p. 265). There have been various forms of such groups, such as sensitivity training (or T-), human relations, or personal growth groups, and they have been applied in therapeutic, personal, professional, and educational settings. These all shared the common goals of seeking personal change through generally non-directed human interactions in groups (Lieberman, Yalom, & Miles, 1973).

The mechanisms that were theorized by Rogers as leading to changing one’s *self* begins with the unconditional positive regard, or prizing, toward the other (Rogers, 1969). Provided with such care and support, people are free to remove their facades and in congruence between their actual experience and their self-picture. Over time, the relationship patterns that people play a part in forming in their everyday lives appear in the life of the group, what is today known as the social microcosm (Yalom & Leszcz, 2005). While a whole range of these patterns are expressed, some of them are maladaptive and impede personal growth (Kiesler, 1996). The encounter group is tasked with exploring these impediments in the context of people’s relationships, so that each participant can learn what feelings their behaviors evoke in others and what responsibility they have in changing their relations. In contrast to conversations that deal with depersonalized knowledge, this relational focus between members of an encounter group is known as the process-focus in the *here-and-now* of the group (Yalom & Leszcz, 2005). The feedback that participants get about their increasingly close relationships to others “appears to be one of the most central, intense, and change-producing aspects of group experience” (Rogers, 1970, p. 33). The changes that people make within the group are later applied to their everyday lives. In this way, the encounter group focuses on shared experiences as a way for people to learn about and intentionally transform themselves.

**Integrating idea- and person-centeredness: The KES framework**

Taking idea-centeredness to be a nuanced, yet essential factor that distinguishes between a type of socioculturally minded learning communities and other learning community approaches, we can identify why adding person-centeredness to idea-centeredness is without precedent. Although learning communities are often focused on the learning of disciplinary content, this does not make them explicitly focused on advancing collective knowledge. From the perspective of knowledge building communities, it is conceivable to see the value of adding encounter groups activities to actively attend to the self and enhance the participants’ learning.

When participants enter into an idea- or person-centered group or community, they focus on two of the following three dimensions: knowledge (K), experiences (E), and self (S). Idea-centeredness focuses on advancing community knowledge as participants share the experiences of working together. The participants’ selves may be important, but in practice the designs only passively attend to them. In contrast, person-centeredness focuses upon self by getting participants to reflect on their experiences and who they are in the ongoing activities. While knowledge may be important, it is only a secondary concern of the design. Thus, having a design that integrates idea- and person-centeredness serves two complementary goals that are linked by the shared experiences of the members in the learning community.

The KES framework (Table 1) shows the relationship between knowledge, experience, and self, spanning the *here-and-now* of the community or from the past and outside of it (*there-and-then*). There-and-then knowledge refers to content previously known (K₁), past or present experiences from students’ everyday lives (E₁), or descriptions of a person’s self outside of the community (S₁). Here-and-now knowledge refers to new content the students are advancing (K₂); current learning experiences (E₂); or a person’s self within the community (S₂).
Table 1: Operationalization of the KES framework

<table>
<thead>
<tr>
<th>Idea-centered design</th>
<th>Person-centered design</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Paragon example: KBCs</em></td>
<td><em>Paragon example: Encounter groups</em></td>
</tr>
<tr>
<td><strong>Knowledge</strong></td>
<td><strong>Experience</strong></td>
</tr>
<tr>
<td>(K₁) Things that people say explicitly, or can be implied from what they say, that reflect the conceptual framework of learning that they had before the learning community (e.g., learning as transmission, individual learning, etc.)</td>
<td>(E₁) Things that people say explicitly about their experiences in life before and outside the learning community or about the context itself (e.g., situations they found themselves in, a characterization of the situation/setting, etc.)</td>
</tr>
<tr>
<td>(K₂) Things that people say explicitly, or can be implied from what they say, that reflect the changing conceptual framework related to the domain that they gained from the learning community</td>
<td>(E₂) Things that people say explicitly about their experience in the learning community or about the learning community itself (e.g., we had a card activity, this is what happened when we…)</td>
</tr>
</tbody>
</table>

**Methods and analysis**

To elucidate the way that learning in a humanistic knowledge building community occurs, we investigated a graduate course in an educational technologies program that had been running for the past decade and whose design explicitly drew upon idea- and person-centered activities. The design was based on the lead instructor’s (third author of this paper) unique background in both knowledge building communities (idea-centered) and sensitivity training groups (person-centered). Specifically, “Challenges and Approaches to Technology-Enhanced Learning and Teaching” (CATELT) had the triple aim of, first, introducing the participants to knowledge about human learning [K]; second, having the students experience the myriad challenges and approaches of technology-enhanced collaborative learning [E]; and third, for students to consider and reflect on themselves as learners [S]. To elucidate how these three goals were interconnected and mutually supportive of the other, we took an ethnographic approach that focused on two specific research questions: (1) How do students’ knowledge, experiences, and self co-develop in humanistic knowledge building communities? (2) What role does the integrated person- and idea-centered design have in fostering these co-developments?

**Design of CATELT**

CATELT was structured as a blended course, where weekly 210 minute face-to-face (ftf) meetings alternated with ongoing activities for the remainder of the week in a wiki environment. Activities were generally designed to promote knowledge advancement through collaborative experiences (idea-centered) or focus on people’s experiences and selves (person-centered). Even though each activity had a different foci, person- and idea-centered artifacts and dialogue occurred in ftf discussions or on the wiki. The continual building-on of previous material between the different spaces as well as over time led to a fluidity of learning that had to do with knowledge, experience, and self over the different activities. Due to space limitations in this conference paper, we refer the reader to several published studies about this unique design (Hod & Ben-Zvi, 2014; 2015).

**Data collection and analysis**

Our data corpus was drawn from a full 13 week semester of CATELT, which included an entire group of 14 students, an instructor (moderator), and a teacher’s assistant (TA). After following all ethical protocols, we collected audio and video recordings of every ftf meeting as well as online artifacts created on the wiki by the students. Throughout the semester, the lead researcher conducted open interviews at opportune times when something interesting occurred and he wanted to know more about it. These data provided a start for a grounded theoretical analysis (Strauss, 1987). Over time, we developed an operational framework for identifying the three types of utterances (knowledge, experience, or self) and two categories for each (there-and-then; here-and-now) (Table 1). These are indicated using brackets in the findings section below. Although we analyzed the full corpus of students using various raters, in this paper we seek to show some of the different patterns we have found within
the full context of one student’s learning. The purpose of this is to be able to have an in-depth discussion about the meaning and usefulness of this framework in understanding the deep transformational changes that can occur in these types of learning environments. Therefore, we carefully analyzed and report on one particular student who showed rich examples of integrations between her knowledge, experience, and self. To ensure that the inferences were reliable, we conducted micro-analysis meetings with a team of researchers who were all familiar with the environment and we triangulated multiple sources of evidence (Schoenfeld, 2007).

A case study of Abby

Abby, a 52-year-old wife and mother of three, entered CATELT with a strong resume of professional experience suggesting she was an intelligent, hard-working, and resourceful learner. Specifically, having earned an M.A. in Chemistry many years before, she had a successful career working as an educational software developer, a project manager at a hi-tech company, and recently as a developer of school-based digital curricular materials in chemistry. She succeeded in what she described as a stressful work environment, where she preferred “logical things, mathematical equations and objective reality” to “wordy theories” (S1). Moreover, Abby was “used to working individually and not as a team” (E1); and she had the perception that she was expected to cover “as much content in the topic that is being studied” (E1). Abby saw learning very much as a depersonalized process, being measured by efficiency and quality, in what can overall be described as a product-orientation.

Abby showed reluctance from the start of the semester about the types of learning activities the community engaged in, as well as about the ideas they were discussing. In addition to her proclivity towards “formulas and logical thinking” (S1), Abby related this type of reflective exercise to her professional colleagues: “They have group-building days… but they are competitive games. I am sure they would really laugh if they saw this” (E1). Likewise, she described the general product-orientation of her work experiences.

During the first week’s introductory presentation about the learning sciences, the moderator discussed traditional models of education that emphasized frontal teaching, coverage, and knowledge transmission. Abby showed her skepticism towards knowledge from the learning sciences by defending traditional school practices:

Ftf 01 GRS: I am assuming that the University is full of researchers who learned in the paradigm of transmission of knowledge. We reached very nice and large achievements. You can’t completely dismiss this [K1].

Abby remained critical of learning community and collaborative approaches to learning over the first several weeks of the CATELT, showing how her work experiences shaped her knowledge of the current experience she was going through.

KES Pattern 1: Re-interpreting prior experience

During the third week, another student, Jihan, was the center of a group reflection session. The conversation focused on the challenges she faced of having to read, write, and speak in a non-native language (Hebrew) as part of the course. Abby—a native Hebrew speaker—appeared to show great empathy with Jihan’s challenge, stating “It is a very important fact that we need to consider… [E2].” Abby related this present experience to her knowledge of learning and the implications this had on herself as an educator. She related this to the ideas of ‘all knowledge builds on prior knowledge’, which the students had been reading and discussing from How People Learn, by considering the role of the teacher:

Ftf 04 GRS: Let’s say a teacher works with students… here we get into the issue of previous knowledge [K2]. They [students] bring something from home, it doesn’t matter from where, they have something that sits in their head. One of our assignments [as teachers] [S2] is to enter their head and to understand why they think that way.

Abby further related this idea to different aspects of her everyday life. For example, she introduced aspects of her work experiences that were not so product-oriented and formulaic, but focused on previous experiences that included humanistic aspects:

Ftf 04 Int Dis: When I worked in a previous job, there was a review every year. They did a personal review of each worker. This review is composed of… first of all you write about yourself. There are about ten sections [details some of them]… Once you finish writing all these things… you pass it to your manager… then you sit and discuss together… It is a process that everyone goes through [E1].
KES Pattern 2: Working through dissonances

During the week six group reflection session (GRS), Abby became the focus of discussion. Consistent with Abby’s product-orientation, she had been experiencing growing discomfort with the unending and expansive activities involved in the way learning was organized in CATELT. She described how these activities related to her personal experiences and self as a learner in the community:

_Ftf 06 GRS:_ I always have the feeling that we are not going in depth into the issues… I feel a type of fluttering on all of the things [E1]. I don’t know, for me personally this is hard because I’m usually very fundamental. I like to go into something, to dive, to understand it until the end, and to feel at the end of the course that I really acquired knowledge… [S1]. And I had this feeling after I saw everything that I needed to do, that I will simply lift up my hands and not do anything [E2].

The moderator continued to inquire about Abby’s statement by connecting her current experiences to her experiences and self outside the learning community, using a “flowers in a field” metaphor that she had used:

_Mod:_ Do you know situations like this where you are in a field and you don’t have time to get to all the flowers? Does this resemble other situations in life?

_Abby:_ I am trying to think. Obviously I got stuck in situations where there were many things and I had to prioritize [E1]. But here I wanted to get to everything, and there wasn’t anything where I said, ‘okay here I don’t have an interest and I don’t want them’. Here I wanted to do all the things, but simply I couldn’t do them all [E2]. This was a disappointment with myself, but no, I know that my time is [trails off] (S2).

_Class:_ [silence for 15 seconds]

_Abby:_ Maybe also losing control [E2]. Meaning, not losing control. I usually plan, and I do the things, and I do everything [E1S1]. And here I didn’t get to everything [E2].

The 15 second silence signified that Abby was realizing something important about herself difficult to articulate. In her work life, Abby knew that she had to prioritize her activities. Yet, in her role as a learner in CATELT, she had trouble reconciling that she needed to be active in an unending learning process. Abby appeared to be at a meaningful juncture where she was understanding that learning was a process. The moderator’s probing put her at a loss for words because her current experiences and the type of characteristics that it required to succeed conflicted with the way she had been describing her prior experience and self.

KES Pattern 3: Interconnecting knowledge and self

Following the sixth ftf meeting, Abby took the lead in creating an elaborate concept map with a peer. The map attempted to show the relation of many of the key ideas from the week’s reading on situated cognition, along with some other ideas from previous readings. Abby volunteered to present the concept map during the seventh ftf meeting. Abby’s reflective diary following the ftf meeting continued on the theme of being product- or process-oriented that she had been concerned with previously (see patterns 1-3). Specifically, she wrote enthusiastically about new knowledge about learning that she now appeared to be making sense of, as related to the concept map, her experiences in CATELT, and her prior experiences:

_07 Online diary:_ In class I presented the concept map that I built because I wanted to get detailed feedback [E2]. One comment that I got: The central idea is not clear. You are right. An additional comment: There is a process here [K2]. Obviously! As someone who worked with chemistry and science, as someone who managed many projects – it is obvious that there is a process…[E1]. When I read the summary of the article where it was written ‘that knowledge is the result of authentic activity that is done in a cultural context’ immediately I saw across my eyes the appropriate concept map to describe the process: there is a result that is the product, and if there is a product, then there is a process and there is a start [K2E12].

After pointing out the process by which she received feedback (E2), Abby mentioned there was a “process here”, referring specifically to what her concept map showed (K2). The use of the exclamation point suggested that this was a large insight for her. Abby then went on to relate this to her personal experiences of
working with chemistry and science (E₁). Her description of the process starting at a certain point and continuing
to evolve was a further elaboration that integrated between the idea of ‘learning as a process’ and her personal
experiences. With this integration between the knowledge of situated learning (as she understood it) and her past
and current experiences, Abby continued to articulate these notions. Directly quoting the article, she described
the process by which she understood its meaning (“immediately I saw across my eyes the appropriate concept
map...”). Although she did not specify whether she was referring to her past or current experiences, based on the
previous statements we can infer that she could have been referring to either one.

KES Pattern 4: Holistic change
By the end of the semester, Abby had made what can be considered to be a transformational change. First, she
was deeply involved in participating in the learning community, despite her initial resistances. Second, on her
own volition, she attended a conference on collaboration following the conclusion of the course, a practice which
she continued regularly after the semester concluded. Finally, following the course, Abby went on to complete
her Master’s thesis, and thereafter enrolled in the Educational Technologies PhD program, examining related
learning phenomena as part of her research.

In a post course interview eight weeks after the course concluded, Abby was asked if she changed during
the course, and if so, to elaborate on how. Her response showed an integration of her knowledge, experience, and
self, spanning both the past and present:

Int-01. …when I come [to this course], I talk, and I say what I feel, and I say what is on my
heart, and this is the thing that is new for me [E₂S₂].

Int-02. Usually I come to a place of work, and I need to work [pounding fist of table], and
nobody really cares what I really feel... What they want from me is usually a product. The
process is less important, exactly the opposite from here [E₁,₂]

Int-03. I now agree that knowledge is built collaboratively [K₂]. I agree that what is important
is the process and not the product itself [K₁,₂]. I agree that you need to strengthen everyone.
That you need to pay attention to everyone [K₂S₂]

Int-04. To talk about myself, okay, to talk about feelings, to talk about thoughts, to talk about
how I learn, how I work, all the metacognition, all of this analysis [E₂] – these are things that I
didn’t have, this is one thing [S₁,₂]. The second thing that really [emphasis] changed for me is
to understand the contribution of the community in the shared learning [K₂]. I didn’t think this
was important at all [K₁]. And today I am found with a feeling that there is a lot to it [K₁,₂].

In Int-01, Abby discussed her experiences in the community in relation to herself as a learner. She then
connected this to her experiences outside the course (Int-02), drawing a comparison between her work experiences
and CATELT. In Int-03, Abby connected this to the ideas studied during the semester, such as knowledge and
collaboration. Inserting her own beliefs (“I agree with”) and values (“you need to...”) showed that these were not
just abstract concepts, but had relevance to the way these ideas applied to herself. Finally, she integrated this
knowledge and experience with her self when discussing her transformation (Int-04: “things that I didn’t have” /
“thing that really changed in me” / “today I am found with a feeling”), by describing her current self as a learner
using new ideas about learning (e.g., metacognition, community, shared learning).

Discussion and conclusion
Our main goal in this article was to explore how knowledge, experience, and self can co-develop in humanistic
knowledge building communities. Sociocultural theory sees the person and activity as irreducible, such that ideas
are always situated and must be viewed holistically (Herrenkohl & Mertl, 2010; Lave & Wenger, 1991). This
study moves the conversation forward by showing how idea- and person-centered designs, each linked to a
different academic heritage, gives attention to different aspects of this holistic learning, and how these span the
there-and-then with the here-and-now. Our assumption is that integrating these two centers within one design can
foster a unique type of learning, not exactly like idea-centered learning communities, but also different from
person-centered groups. The findings in this study provide different insights into the unique ways this happens.

KES Pattern 1 shows the subtle way that the new experiences and knowledge gained from participating
in a community can be applied back into a person’s life, leading them to re-interpret past events. In Abby’s case,
after she had new experiences in the learning community she began to bring up past relevant material from her
life that she had hitherto left out. The knowledge which Abby gained provided a new cognitive framework that
allowed Abby to reconsider her past experiences and self. Her ability to do this appeared to be a breakthrough for her, now being able to ‘fit in’ her new knowledge in a way that was consistent with her previous experiences.

KES Pattern 2 shows how knowledge, experience, and self are constantly negotiated between the there-and-then and here-and-now, and how dissonances between these can lead to significant insights. In Abby’s case, moderated reflection led her to reconsider lifelong practices that brought a usually articulate person to silent rumination. For Abby, this moment was critical in her transformation. The 15 seconds of silence and her incoherent remarks afterwards had reached a point where they were publicly inconsistent with each other, brought to light by the moderator’s questioning. Although it took six weeks of sustained reflection to get to this point, findings 1 and 2 show how the seeds grew to a point where she had to make many interconnections between her knowledge, experience, and self in a way that bridged the past and present.

KES Pattern 3 shows the delicate and subtle nature of knowledge, and how deeply contextualized it is not just within a person’s prior knowledge, but with their experiences and self. Over the course of a semester, Abby learned to understand ideas like ‘learning as a process’ or ‘collaborative learning’ in new ways. Taking a purely cognitive perspective could easily lead to thinking that Abby learned little—just a few concepts—over 13 weeks. It could lead to criticism of Abby that her interpretation of situated cognition largely missed the point. By understanding how Abby’s new knowledge was tied to her experiences and self, the insight that Abby made could be valued for how truly transformative it was. It is possible to appreciate how a person can miss the canonical points about a topic (in this case situated cognition, which is much more than ‘learning is a process’), yet what she did take from the topic was still highly important and significant within her own private logic.

Finally, KES Pattern 4 showed how Abby’s transformation within the humanistic knowledge building community was deeply related to her co-developments in knowledge, experience, and self. Throughout the semester, she was given the opportunity to fully participate by sharing her knowledge, experience, and self with others in the community, learning by comparing what others brought to her own experience, and reflecting on deep questions about herself. While the course was only 13 weeks, it ultimately helped Abby make a transformation such that she changed lifelong learning practices despite her strong, initial reluctance.

By considering these patterns together, we are able to see the importance of both the idea- and person-centered aspects of the design, and the way they support one another. Had CATELT not allowed students to discuss their knowledge, experiences, and self in an ongoing, intensive, and deliberate way throughout the semester, it is likely that Abby would not have made such a holistic transformation. Indeed, we argue that because of the integration of the two centers, she had opportunities to sort out these complex ideas as they were relevant to her life. For example, the climate fostered through person-centeredness (care, prizing, and empathic understanding) allowed Abby to draw out her prior knowledge, experiences, and self that later shaped her knowledge about learning. Or, the new knowledge that Abby gained about learning contributed ultimately to her belief that the process, and not just the product, is important in learning. The mutuality of the findings suggests a view of learning that is much less linear, and much more like pieces of a puzzle fitting together. Even in this analysis, the borders that we draw between the findings are somewhat arbitrary. They do represent significant episodes of learning within CATELT, but they are accumulative and touch on the broader narrative of Abby’s life holistically. To sum, our analysis of Abby suggests that knowledge, experience, and self are all interrelated between the there-and-then and here-and-now. When provided with sufficient opportunities and support to explore these, as humanistic knowledge building communities aim to do, the related domains of knowledge, experience, and self can co-develop and ultimately lead to transformational change.

Implications for theory and practice
The key practical implications of this research in both K-12 and higher education run in two directions. For knowledge building communities, it is in the prominence given to encounter group types of activities within the course design. Even in learning communities that have person-centered activities, ours is a highly unusual approach in terms of the amount of time allocated for these activities. The purpose of giving so much time for the person-centeredness is based on our serious belief in the educational principles of Carl Rogers, where the person is accepted unconditionally and is given a fertile space to grow through their exploration of self through the other. In our unique design, person-centeredness is not just something that occurs towards the start of the semester, or at the end of a course meeting when there is time left to reflect on thinking. Person-centeredness is given prominence throughout the semester, with activities designed to both maintain and continue deepening students’ relationships and trust. It is in this person-centered way that we view the learning community, and why we feel it is necessary and appropriate to distinguish it from other knowledge building communities with this label.

Conversely, this research can contribute to person-centered approaches by showing the importance of idea-centered designs. One of the large and perennial criticisms of person-centeredness, akin to discovery learning or student-centered learning, is the lack of content or knowledge goals. Students in person-centered designs may
learn what interests them, but they may lack a true appreciation for what it means to fully engage in a disciplinary practice. By bringing in idea-centered approaches, students have the opportunity to explore what doing knowledge building in a particular discipline means. Within this context, they are then able to consider their prior experiences and selves and make decisions about whether these new experiences are those they value.

While designing and implementing a humanistic knowledge building community is a highly attainable goal, this comes with risks and the need for highly skilled, professional moderation. Making time for person-centeredness means allowing for unanticipated issues that the students raise, such as resistances towards the moderator, dealing with interpersonal conflicts, and other issues long described in psychological literature (e.g., Bion, 1959). It is completely understandable why teachers prefer to put aside these social and emotional matters, or deal with them individually and not as a group level process which they nurture. Moderating encounter groups often requires licensing, and the risks of running these groups can potentially disrupt students’ lives if not treated with the professional sensitivities necessary. In general, we think there are ethical consideration that must be taken into account for anyone leading such a community. With that said, we do believe that fostering transformational change requires learning environments that are sensitive to the student as a whole.

Considering idea- and person-centeredness as two foci with different goals and activities helps shows the relation between what may appear to be otherwise fragmented lines of research. We have been engaged in a long-term design experiment, and although we have published different aspects of it, this is our first offering of an overarching framework to explain a great deal of our efforts. Our research has now matured to the point where we want to share our own conceptualization and findings with the broader educational community. Mainly, we seek to show how the integration of idea- and person-centeredness—embodied in what we call ‘humanistic knowledge building communities’—can inform the theory and practice of learning communities. We believe that the impact of this research has the potential to start a conversation between two academic communities, defined by idea- and person-centeredness, that have yet to converge.

References
Herrenkohl, L. R., & Mertl, V. (2010). How students come to be, know, and do: A case for a broad view of learning: Cambridge University Press.
A Principled Approach to Designing Assessments That Integrate Science and Computational Thinking

Satabdi Basu, SRI International, satabdi.basu@sri.com
Kevin W. McElhaney, SRI International, kevin.mcelhaney@sri.com
Shuchi Grover, shuchig@cs.stanford.edu
Christopher J. Harris, SRI International, christopher.harris@sri.com
Gautam Biswas, Vanderbilt University, gautam.biswas@vanderbilt.edu

Abstract: There is increasing interest in broadening participation in computational thinking (CT) by integrating CT into precollege STEM curricula and instruction. Science, in particular, is emerging as an important discipline to support integrated learning. This highlights the need for carefully designed assessments targeting the integration of science and CT to help teachers and researchers gauge students’ proficiency with integrating the disciplines. We describe a principled design process to develop assessment tasks and rubrics that integrate concepts and practices across science, CT, and computational modeling. We conducted a pilot study with 10 high school students who responded to integrative assessment tasks as part of a physics-based computational modeling unit. Our findings indicate that the tasks and rubrics successfully elicit both Physics and CT constructs while distinguishing important aspects of proficiency related to the two disciplines. This work illustrates the promise of using such assessments formatively in integrated STEM and computing learning contexts.

Introduction

Driven by the needs of a 21st century workforce, education and industry stakeholders recognize that computing knowledge and skills provide the foundation for competency in a multitude of fields (Wing, 2006). One approach for making computational thinking (CT) accessible to K-12 (e.g. formal precollege) students is to integrate it with existing components of the K-12 Science, Technology, Engineering, and Mathematics (STEM) curricula. STEM topics lend themselves particularly well to integration with CT, because many of the epistemic and representational practices central to expertise in STEM disciplines (e.g., characterizing problems and designing solutions, developing and using models, analyzing and interpreting data) are also primary components of CT proficiency (Basu et al., 2016). The integration of STEM and CT in K-12 settings is further motivated by current STEM workforce practices that increasingly rely on computational modeling and simulation tools for understanding, analyzing, and solving problems (Landau, 2006; Freeman et. al., 2014). The US Framework for K-12 Science Education (NRC, 2012) and Next Generation Science Standards (NGSS Lead States, 2013) instantiate this view by including ‘Using Mathematics and CT’ as a key science and engineering practice.

With computer science frameworks (e.g., the US K-12 CS Framework, 2016) gaining traction in K-12 instructional settings, science will likely emerge as an important context for teaching CT in school. Leveraging the synergy between CT and science in K-12 classrooms will require, among other things, the systematic development of assessments that measure learning at the intersection of science and CT. These assessments will need not only to integrate the science and CT disciplines, but also integrate disciplinary concepts and practices, following the vision put forth by contemporary STEM education frameworks that integrate content and practice.

In this paper, we describe a general principled approach for designing rich assessment tasks and associated rubrics that integrate science disciplinary knowledge, CT concepts, and computational modeling practices using Evidence Centered Design (ECD) principles (Mislevy & Haertel, 2006). We discuss an application of this approach where we designed and administered multiple assessment tasks embedded within a web-based, computational modeling environment that supports the integrated learning of physics and CT for high school students. Using one such task and an associated rubric as an example, we use video recordings of students responding to the task to analyze students’ responses to the task, illustrating (1) how the task elicits different aspects of students’ science (physics) and CT proficiencies in this integrated domain and (2) how the rubric distinguishes these aspects of proficiency for the purposes of formative assessment.

Synergistic learning of science and CT

Developing a computational model of a physical phenomenon involves integrating key aspects of CT and scientific practice: identifying appropriate abstractions (e.g., underlying rules governing the behavior of relevant entities), making iterative comparisons of the generated representations with the target phenomenon, and debugging the abstractions to generate progressively sophisticated explanations of the phenomenon.
research studies have shown that integrating CT and scientific modeling can be beneficial (e.g., Hambrusch et al., 2009; Blikstein & Wilensky, 2009; Basu, Biswas & Kinnebrew, 2017). Sengupta et. al (2013) describe how integrating CT and scientific modeling can be beneficial: (1) Lower the learning threshold for science concepts by reorganizing them around intuitive computational mechanisms: computational representations introduce discrete and qualitative forms of fundamental laws, which are simpler to understand than equation-based continuous forms (Redish & Wilson, 1993); (2) Programming and computational modeling as representations of core scientific practices: Soloway (1993) argued that learning to program amounts to learning how to construct mechanisms and explanations; and (3) Contextualized representations make it easier to learn programming (Papert, 1991). These benefits reflect the framing of proficiency in both science and CT (by the NGSS and K-12 CS Framework, respectively) as the integration of knowledge and practice.

Evidence-Centered Design
We use ECD, a principled assessment design framework (Mislevy & Haertel, 2006), to create assessments that are inclusive of science and CT concepts and practices. ECD promotes coherence in the design of assessment tasks and rubrics and the interpretation of students’ performances by explicitly linking claims about student learning, evidence from student work products, and design features of tasks that elicit the desired evidence. ECD begins with a domain analysis, which entails gathering and organizing information on the domain to be assessed. This is followed by domain modeling, which entails the articulation of specific learning targets and task design specifications, which in turn inform the development of tasks and rubrics. ECD has been used to develop CS assessments for the Exploring Computer Science curriculum (Goode et al., 2012), as well as to develop science assessments that integrate content knowledge with science practices along the performance dimensions of the NGSS for summative and formative purposes (Harris et al., 2016).

Methods
Designing synergistic assessment tasks for measuring science and CT proficiencies
Figure 1 illustrates an ECD process for creating assessments that are inclusive of science and CT, while targeting concepts and practices that cut across science and CT. Our process begins with identifying the integrated science and CT domain, for example ‘High school kinematics and CT’, or ‘Middle school carbon cycle and CT’. Then, in the domain analysis phase, we unpack the three domains of science disciplinary concepts, CT concepts, and computational modeling practices. We elaborate on and document the target constructs that we want to assess in each domain, determine assessment boundaries and expected background knowledge for the domains, and articulate the knowledge, skills, and abilities (KSAs) relevant to each domain. Next, we create integrated domain maps to represent the relationships and synergies between the three domains. These maps are important because they enable us to be principled in our choice of which science and CT concepts and modeling practices to integrate in an assessment task. The integrated domain maps offer a range of ways to coherently express integrated learning goals for science and CT during the domain modeling phase.

The integrated learning goals constitute the claims we make about what students should know and be able to do. Additionally, for each learning goal, we articulate a design specification that guides the design of tasks and rubrics aligned to it (Mislevy & Haertel, 2006). Each design specification focuses on the following aspects that provide the basis for tightly integrating task and rubric design: (1) focal KSAs, (2) features of student responses that constitute evidence of proficiency with each focal KSA, (3) characteristic features of assessment tasks that can effectively elicit this evidence of proficiency, and (4) variable task features that can shift the difficulty or focus of a task. In the task and rubric development phase, these design specifications and technology affordances of the task delivery system inform the development of tasks and rubrics in a way that aligns the assessment targets, desired evidence of student proficiency, task design features, and scoring criteria. Though the design process may appear to be linear and unidirectional, it is iterative in nature, allowing developed tasks to help refine the learning goals or design specifications, for example.

Applying the approach described in Figure 1, we have developed multiple assessment tasks to measure high school students’ integrated proficiencies in Physics (high school kinematics) and CT. An initial domain analysis helped us identify a set of target constructs for each of the domains of physics disciplinary concepts, CT concepts and computational modeling practices. Based on these constructs, we articulated a set of learning goals, each integrating physics concepts, CT concepts, and an aspect of computational modeling practice.

Figure 2 illustrates selected constructs that we identified in each domain, as well as a few sample learning goals that we articulated by integrating the constructs. For example, the first learning goal articulated in Figure 2 (in boldface) integrates target constructs from all three domains (in boldface). Before articulating learning goals, we created integrated domain maps where we identified key relationships between the physics and CT domains.
to ensure that the integration of the physics and CT concepts for each learning goal leveraged the synergy between the domains (instead of combining physics and CT concepts arbitrarily). For example, calculating the velocity of an object based on its initial velocity, acceleration and time closely relates to the CT concepts of initializing and updating variables (velocity, acceleration and time are all examples of variables), and operators and expressions. Additionally, combining the related physics and CT concepts with different aspects of computational modeling practices like ‘Develop, Use, Test, Debug’ helped create learning goals that guided task design specifications at different levels of complexity.

Based on the learning goals, we developed 18 tasks of varying complexity comprising various formats such as multiple choice, explanation, and programming. In some tasks, we provided most of the code and asked students to fill in a small part that targeted a specific concept, while in other tasks, we provided required blocks and asked students to focus only on arranging the blocks in a correct computational sequence. We created different versions of debugging tasks such as asking students to correct a given buggy program; showing students a snapshot of a program and asking them to indicate which block(s) to modify and how; and asking students to use resultant data and graphs to identify errors in a hypothetical program not shown to them.

Figure 1. Design process schematic for assessment tasks that integrate science and CT learning.

Figure 2. Unpacking the physics and CT domains, identifying their relationships through integrated domain maps, and the articulation of integrated learning goals. (Bold text illustrates how a learning goal integrates concepts and practices across disciplines).
Empirical study using assessments to elicit integrated science and CT proficiencies
We embedded the assessment tasks in the C2STEM learning environment – a browser-based system that engages students in computational modeling and simulation of Physics phenomena. The computational modeling representation uses custom domain-specific blocks developed on top of NetsBlox (Broll et. al., 2016), a block-based extension of Snap! (http://snap.berkeley.edu/) to help learners focus on physics concepts.

We conducted an empirical pilot study to examine how well our assessment tasks elicited students’ proficiencies in integrating physics and CT, and how rubrics could be designed to distinguish between components of proficiency across students. The study was conducted within a high school summer program for Science and Math. The students worked on three C2STEM modules as part of a 10-hour kinematics curriculum, with each module comprising an alternating sequence of scaffolded modeling activities and embedded assessments. All of the participating high school students had prior experience working with NetsBlox as part of prior summer school activities, and some reported familiarity with languages like Scratch and Python, but none had taken a high school physics class.

In this paper, we limit our analyses to one assessment task, the Airport task (Figure 3), which addresses the learning goal ‘Develop a computational model that simulates 1-D, constant velocity motion using addition of velocity vectors that occur only under particular conditions.’ We examine 10 students’ responses (4 female, 6 male) to this task to determine how well it elicits evidence for target physics and CT constructs in the context of computational modeling, and also differentiates among levels of proficiency within the domains.

Figure 3. The “Airport Task: An example programming assessment task.

Data sources and plans for analyses
We recorded all student responses to assessment tasks using the Camtasia™ screen-capture software. We examined the screen recordings to characterize students’ model-building approaches and challenges faced. We noted whether students solved the tasks correctly on their first attempts or whether they required multiple iterations of testing and debugging. For students submitting an incorrect solution, we recorded the different types of errors and verified that the students made an honest attempt to solve the tasks. Based on the analysis, we
developed a rubric (Table 1) that scores students’ final programming solutions (not their model-building approaches) along two aspects of integrated physics-CT proficiency: (1) the ability to express physics relations in a computational model and (2) the ability to use programming concepts to model a physics phenomenon. Scoring the task based on these distinct aspects of proficiency has the potential to provide useful information to researchers and teachers on the specific nature of students’ proficiencies.

Table 1: Rubrics for characterizing student performance on an integrative assessment task

<table>
<thead>
<tr>
<th>Criterion #</th>
<th>Rubric for scoring the Airport task</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Program expresses correct relations among velocity, position and time, and correct units for each</td>
<td>1 point</td>
</tr>
<tr>
<td>2</td>
<td>Program reflects that walking on the moving walkway causes resultant speed to be additive in the x direction (walking speed + walkway speed) and constant (no acceleration)</td>
<td>1 point</td>
</tr>
<tr>
<td>3</td>
<td>Program makes the distinction between actions that need to happen once during initialization and actions that need to be repeated in the simulation step</td>
<td>1 point</td>
</tr>
<tr>
<td>4</td>
<td>Program correctly determines which action always happens and which happens under certain conditions</td>
<td>1 point</td>
</tr>
<tr>
<td>5</td>
<td>Program updates the variable corresponding to Josh’s velocity on the walkway a. under the correct conditions (Use conditionals with appropriate expressions to update Josh’s velocity under correct conditions between Point B and Point C only), and b. in the correct fashion (the x velocity is set to a new constant value instead of changing at every simulation step)</td>
<td>1 point</td>
</tr>
<tr>
<td>6</td>
<td>All code in the program is reachable and can be executed</td>
<td>1 point</td>
</tr>
</tbody>
</table>

Findings

Scoring students’ final programming artifacts for the Airport task using the rubric described in Table 1 revealed that half the students (s1 through s5) solved the task correctly, earning the maximum scores (2 and 4, respectively) on the physics and CT rubric components. Among the students who were unable to solve the task correctly, some students (s6 – s8) demonstrated high proficiency with the physics component (scoring 2 points), but only partial proficiency on the CT component (scoring less than 4 points). Two others (s9 and s10) demonstrated partial proficiency on both the physics and CT components (scoring 1 point on the physics component and less than 4 points on the CT component). We define high proficiency on a component as scoring the maximum possible points on the rubric for the component. Figure 4 summarizes students’ scores.

Based on students’ proficiencies on the two rubric components, we grouped them into three categories – High Physics-High CT, High Physics-Partial CT, Partial Physics-Partial CT. In our small sample, we did not find any student work that we could categorize as Partial Physics-High CT. Next, we discuss example solutions and some of students’ programming behaviors for each of the three student categories.
Category 1: High Physics, High CT: Figure 5 illustrates two correct solutions where the only change students made to the given code was modifying the procedure ‘set-Josh-resultant-velocity’ to specify Josh’s new velocity beyond Point B. In the solution to the left, the student correctly specifies Josh’s velocity as the sum of the walkway speed and Josh’s speed in the ‘else’ part of the given conditional block. In the second solution, the student hardcodes the value of the variable ‘Josh’s speed’ to 2.5 instead of a more general expression. In both examples, the students do not modify the other procedure ‘update-position’, thus maintaining correct relations among position, velocity, and time. All parts of the codes are reachable, and the students correctly distinguish between initialization actions that must occur when the green flag is clicked, and actions that must repeat at every simulation step. These programs meet all six criteria across both rubric components.

While all five students in this category finally produced programs that demonstrated high proficiency in physics and CT, we observed that their pathways to reach the final state varied. Two of the students reached the correct solution on their first attempt, requiring only a single test of the modified program. The three other students initially modified Josh’s velocity before Point B (instead of after Point B) and specified a non-zero y-component of Josh’s velocity. However, they were able to rapidly identify the errors and debug their programs.

Category 2: High Physics, Partial CT: Figure 6 illustrates one of the three student solutions in this category. The student correctly expresses the relations among velocity, position, and time, and correctly expresses Josh’s resultant velocity as the sum of ‘walkway speed’ and ‘Josh’s speed.’ However, the student incorrectly specifies conditions for updating Josh’s velocity by using all points to the left of Point C instead of only points between points B and C (rubric criterion 5). The incorrect conditional statement is the reason for the program earning less than the maximum score on the programming concepts rubric component.

Based on the videos, we observed that all three students in this category went through multiple iterations of testing and subsequent program modification to reach their final program state, confirming that they made legitimate efforts to reach a correct solution. None of the three students had difficulty creating the physics expression ‘Josh’s speed + Walkway speed’, but they made three general types of programming errors related to the CT constructs of variables, conditionals, and control structures. First, students sometimes assigned the physics
expression to an incorrect variable. Second, students specified incorrect conditions under which the expression applies. Third, students used ‘forever’ loops that do not terminate. These examples illustrate ways that students’ proficiency with a physics and CT concepts are distinct.

Category 3: Partial Physics, Partial CT: Figure 7 illustrates one of the two solutions in this category. The student demonstrated an incomplete understanding of both the physics relations and the programming concepts, scoring 1 and 2 points respectively on the Physics and CT components of the rubric. On the physics component, the student incorrectly expresses Josh’s velocity beyond Point B as the product of time and speed (rubric criterion 1). Also, in the procedure ‘set Josh resultant velocity’, the student has incorrectly set Josh’s velocity beyond Point B to zero (rubric criterion 5), thereby updating Josh’s velocity differently in two places in the code. Moreover, the solution incorrectly contains a ‘forever’ loop inside the simulation step, effectively stopping the execution of other code for all objects (sprites) (rubric criterion 6).

From the videos, we observed that the programming behavior and challenges faced by students in this category were generally similar to that of students in Category 2, except that these students were unable to correct either their Physics related errors or errors from incorrect programming constructs. In fact, the physics-related challenges appeared to be compounded by computational challenges. For example, when a ‘forever’ loop in the simulation step for ‘Jos h’ effectively stopped execution of code for other objects (sprites), one student was compelled to modify code for a different sprite (Kate) to model its motion correctly.

Discussion and future work
The recent focus on “CSForAll” (Barnes, 2017) and the policy attention to STEM learning has led to an escalated interest in finding ways to tap into the synergy between CT and science. Making STEM+CT learning successful in precollege settings requires systematically designed assessments for this integrative domain. This paper discusses an approach for designing assessment tasks that target integrated proficiencies across science and CT disciplines, while also differentiating among levels and the nature of proficiencies in the disciplines. The approach has the potential to be generalized to all grade levels and science disciplines. ECD enables us to use a principled approach for assessment development that integrates concepts and practices in the domains while aligning with established education frameworks that integrate content and practice.

Our examination of students’ responses on one such integrated assessment task during a recent pilot study reveal varied physics and CT related challenges that students face while working on such integrative tasks. The ability to identify these challenges can provide valuable information to help teachers guide individual students appropriately (e.g., science disciplinary content versus programming concepts). Our work illustrates the potential value of using such assessments for formative purposes, so that students can achieve synergistic learning of science disciplinary concepts, CT concepts, and computational modeling practices. In order to be useful for formative purposes, assessments must be able to isolate evidence on a specific set of constructs and should not involve additional construct irrelevant activity. Also, varying the task design formats for the same target constructs can help elicit evidence of proficiency at different levels of granularity and provide a more comprehensive assessment of students’ proficiencies.

As future work, we will analyze student responses to integrative assessment tasks from a larger classroom study. We plan to analyze responses to a range of tasks from the kinematics domain and a different physics domain (force) created using the ECD-based approach described above, allowing us to generalize our two-component
rubric framework. Analyzing student work across the two domains will enable us to investigate how students’ CT proficiencies change over time and whether they transfer across domains. Further, we will explore ways to apply the rubrics to observable evidence from log files to facilitate automated scoring of these integrative assessments. Manually scoring students’ programming artifacts using multi-point rubrics requires going through each students’ code and can be labor intensive. While automating the scoring of open ended programming tasks can be extremely challenging, a principled design for focusing on specific constructs in our carefully designed assessment tasks constrains possible student choices and makes automated scoring feasible. Automated scoring will offer opportunities to provide students with carefully designed guidance in real time and provide rapid insights to teachers about their students’ proficiencies that can, in turn, inform teachers’ instructional decisions.

References

Acknowledgments
We thank Nicole Hutchins, Miklos Maroti, Luke Conlin, Kristen P. Blair, Doris Chin, Jill Denner, and our other collaborators from Vanderbilt University and Stanford University for their numerous contributions. This research is supported by NSF grant #1640199.
Investigating the Coupling of Narrative and Locality in Augmented Reality Educational Activities: Effects on Students’ Immersion and Learning Gains

Yiannis Georgiou and Eleni A. Kyza
Ioannis.Georgiou@cut.ac.cy, Eleni.Kyza@cut.ac.cy
Media, Cognition and Learning Research Group, Cyprus University of Technology

Abstract: The popularity of Augmented Reality (AR) technologies is on the rise, yet little is known about the design principles which could support immersion and learning in such contexts. This study investigated the hypothesis that greater coupling between the physical space and the narrative of the AR activity leads to enhanced immersion and learning. Forty-five middle school students participated in this study: students in Condition1 ($n=22$) participated in an AR activity with strong coupling between the narrative and the physical space, while students in Condition2 ($n=23$) participated in a loose coupling version of the activity. Data collection included baseline data, questionnaires investigating students’ immersion and learning gains, as well as post-activity interviews. Findings showed higher learning gains and increased immersion for the students participating in the strong coupling condition than for the students in the loose coupling condition.

Introduction and theoretical background
Narrative-based AR activities have been argued to be effective in engaging students in science learning due to their immersive affordances (Dunleavy et al. 2009). Conceptualizing immersion as a gradated process of cognitive and emotional involvement, researchers have argued that feelings of presence (feeling surrounded by a blended yet realistic physical/virtual environment), and flow (being fully absorbed in the activity), are considered as indicators of full immersion and can be powerful contributors to learning (Cheng et al., 2015). Previous studies have provided empirical evidence arguing that such heightened levels of immersion in narrative-based AR activities can increase students’ performance and subsequent science learning (e.g., Georgiou & Kyza, 2017a). Despite such reports, little is yet known about design principles which can contribute to more engaging AR learning activities, inducing higher levels of immersion.

Reid, Cater, Fleuriot, and Hull (2005a) were the first ones who argued that the coupling between physical space and narrative may affect the external realism of the narrative resulting in different levels of immersion in mobile-based activities. Reid et al. (2005a) distinguished between three “levels of significance to a place”: (a) arbitrary linkage, (b) physicality, and (c) particular location. At the first level, there is no connection between the physical space and the mediated experience. At the Physicality level “there are certain features of a place that are significant, but the actual geography is not important”, and at the “Particular location” level “the actual location and physical artefacts in a place are significant and meaningful” to the experience (p. 28). Based on this distinction, Karapanos Barreto, Nisi, and Niforatos (2012) have proposed that, from a design-based perspective, this may result in (a) mobile-based activities with a strong coupling, as they are enacted at the original location that is directly related to the narrative, and (b) mobile-based activities with a loose coupling, as they are enacted at a location that simply induces the same atmosphere with the narrative.

Existing studies, primarily from the Human-Computer Interaction field, yield inconclusive findings for the coupling phenomenon. Reid, Hull, Cater, and Fleuriot (2005b) investigated users’ immersive experiences in “Riot!1831”, a situated-based interactive drama employing a location-aware, mobile application. According to their findings, the users’ immersion was highly increased during the moments when there was strong coupling between the events in the narrative plot, as it unfolded in the virtual space, and the physical world. According to Reid et al. (2005), during moments of high immersion “the senses are heightened to the coincidence of the event and it feels almost supernatural, as if events in the virtual world have somehow moved across into the physical world” (p. 291). Karapanos et al. (2012) compared the degree of immersion and presence of location-based narratives in different settings and found that the condition with the highest coupling between the place and the narrative resulted in a more immersive experience. However, they did not find statistically significant differences in regard to feelings of presence. Karapanos et al. (2012) explained this latter finding by discussing the limitations of their study, such as the use of a non-standardized scale with a single-item for measuring presence, or the short duration (3 minutes) of the narrative-based activity, which may have contributed to the findings of their study. In a more recent study, Rossitto, Barkhuus and Engström (2016) presented an evaluation of a location-based AR drama activity. Even though the narrative of the activity was loosely-coupled with the physical location where it
unfolded, the users were able to imagine possible ways for interweaving the story with the place. Based on these findings, Rossito et al. (2016) concluded that this process of interpretation and sense-making between the narrative and the space, resulted in a highly immersive experience in this loosely-coupled activity.

Taking into account that existing evidence regarding how coupling influences immersion is inconclusive, the present study compared the effects of two types of coupling (strong/loose) and sought to answer the following research questions: (a) Does the level of semantic coupling influence how students experience immersion? (b) Does the level of semantic coupling have a different impact on what students learn? (c) What are the main factors affecting students’ immersion, in terms of flow and presence, for each type of semantic coupling?

Methodology

This study adopted an experimental research design, which included two groups of middle school age students. Each group of students was randomly assigned to one of two semantic coupling conditions: (Condition1: Strongly-coupled AR educational activity, Condition2: Loosely-coupled AR educational activity). A cross-comparison method was employed for investigating students’ immersion and learning gains in each condition.

Participants

Fifty-five middle school students, who were enrolled in a summer camp at a public university, participated in this study. Data from ten students were removed from the corpus as the students had missed key activities, resulting to a final sample of forty-five students (n=45). Students were randomly assigned to each condition. Condition1 had 22 students (13 boys, 59.1%) and Condition2 had 23 students (12 boys, 52.2%). The mean age in Condition1 was 12.3 years, and the mean age in Condition2 was 11.7 years; there was no statistically significant difference in students’ age between the two conditions (Z=-1.50, p=.135) nor in the baseline assessment (described later).

Intervention

Students in both conditions used the “Mysterious disease” marker-based AR activity; the activity was based on Quick Response (QR) codes and took place at one of the university buildings. The AR activity was a narrative-driven investigation running on Android tablets, and its duration was 45 minutes. Students were randomly divided in pairs and were assigned the task to collaboratively investigate a case regarding the outbreak of a mysterious disease at the university. The pedagogical goal of the AR activity was to engage students in an explanation-building process about the problem-based case, and support students’ understanding of scientific concepts related to the nature of microbes (e.g. types, characteristics, etc.) and the transmission of pathogens. Upon the activation of each QR code, students had access to video-based characters (e.g., university nurse, professor of microbiology, health services agent), who provided them with access to information on different aspects of the learning topic, including relevant evidence. According to the narrative, the mysterious disease was caused by a massive food poisoning due to insufficient adherence to food safety procedures at the university cafeteria. Since the goal of this study was to investigate the impact of loosely-coupled and strongly-coupled AR activities on students’ immersion and subsequent learning, two different versions of the “Mysterious disease” AR activity were developed: a strongly-coupled version and a loosely-coupled version; students in each condition participated in a different version of the activity.

The strongly-coupled AR activity

The strongly-coupled version of the activity explicitly intended to increase the relation of the content in the mobile environment and the physical space in two ways (Figures 1 and 2).

Figure 1. Students participating in the strongly-coupled activity collect data from a real notepad. Figure 2. Students participating in the strongly-coupled activity in the kitchen of the cafeteria.
First, the QR codes were placed at specific points of interest that were tightly connected to the narrative (e.g., at the university’s medical center, where the nurses recorded the symptoms of the affected students; at the university cafeteria, where the cooks did not follow food safety rules; at a lecture hall, where a professor lectured about microbes, etc.). In this version, the path that students walked along was designed to fit the theme and the narrative structure. Second, the places where the QR codes were activated, were also enriched with physical cues which were also aligned with the narrative (e.g., broken freezer, kitchen disposable gloves, printed posters and diagrams, symptom notebook, fermented food whose production uses microbes such as beer, yogurt, etc.). Such designs of the strongly-coupled activity have been argued to support strong blending between the narrative and the physical space by creating “natural coincidences” between events in the virtual and the physical world (Reid et al., 2005b).

The loosely-coupled AR activity
The loosely-coupled version of the activity took place at the same location (at the university premises) where the narrative unfolded (Figures 3 and 4).

Data collection and analysis
This study employed a mixed methods approach. The collected data included baseline data, questionnaires investigating students’ immersion and learning gains, as well as post-activity interviews with 14 students from Condition1 (63.6%) and 14 students from Condition2 (60.9%).

Baseline data
We collected baseline data using a survey, aiming at creating a profile for the students and establishing the equivalency of the two conditions. The survey had two main parts: Interest in science and Students’ attitudes towards tablets. Interest in science was measured using a Likert scale with 10 items derived from the Test of Science-Related Attitudes (Fraser, 1981), as this was adapted and validated in the study of Bressler and Bodzin (2013). The Cronbach’s a for the adapted instrument was 0.90. Students’ attitudes towards tablets was assessed using the Computer Attitude Measure for Young Students (CAMYS, Teo & Noyes, 2008), which was composed of 12 items, using a five-point Likert scale. The CAMYS is considered a valid instrument and has a documented reliability alpha coefficient of .85. Differences between the two conditions were examined using the Mann-Whitney U test, given the small sample size of each condition and the lack of normal distribution in the data.

Immersion and conceptual gains questionnaires
After the implementation of the “Mysterious disease” AR activity, the students in each condition completed the Augmented Reality Immersion (ARI) questionnaire (Georgiou & Kyza, 2017b), which measured their experienced immersion, and a conceptual assessment test to investigate learning gains. First, the students were asked to complete the Total Immersion scale, which is comprised of two subscales with a total of seven items: Flow (3 items) and Presence (4 items). The Cronbach’s a for the Total Immersion scale has a documented reliability alpha
coefficient of 0.82, while the Cronbach’s α for the Flow and the Presence subscales are .87 and .80 correspondingly. Second, a conceptual understanding test was administered to assess the differences in students’ conceptual learning about the scientific concepts related to topic of the AR investigation. The conceptual understanding test was composed of ten multiple-choice items and five open-ended questions. The aim of the multiple-choice items was to evaluate students’ factual knowledge, while the aim of the open-ended questions was to evaluate students’ reasoning about microbes and the transmission of pathogens. A Mann-Whitney U test was used to investigate the differences between the two conditions, given the small sample size of each condition and the lack of normal distribution in the data.

**Post-activity interviews**

Fourteen students from each condition participated in semi-structured individual interviews, which took place after the intervention. The average time of the interviews was 19.09 mins (SD=4.34 mins). Students were initially asked to report feelings associated with their experience during the immersive AR activity, and then were probed to discuss feelings of presence (e.g., *To what extent did you experience this problem-based investigation as a fictional case? Did you perceive the narrative-based storyline as a real one or not?*), their feelings of flow (e.g., *How focused did you feel during the problem-based investigation? To what degree did this investigation manage to capture all of your senses?*), as well as the factors which had affected these feelings positively or negatively (e.g., *What were the main factors affecting your feelings? How did these factors affect your feelings?*).

All interviews were transcribed and qualitatively analyzed using a thematic analysis approach (Attride-Stirling, 2001), to investigate the factors students reported as affecting their feelings of presence and flow. This analysis led to the identification of basic thematic categories, and factors discussed within each thematic category which were hypothesized to influence the feelings of presence and flow during the AR activity in each condition. Each unique factor was counted once for each student (absence/presence) and was coded as negative or positive, thus providing a more nuanced indication of how each factor affected students’ immersion.

**Findings**

**Setting the baseline**

A Mann-Whitney U test was used to identify any potential differences between students in the two conditions, in terms of their prior interest toward learning science and attitudes towards tablets use. As shown in Table 1, there were no statistical differences between the students in both conditions.

<table>
<thead>
<tr>
<th></th>
<th>Condition 1 (Strongly-coupled AR)</th>
<th>Condition 2 (Loosely-coupled AR)</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Interest in Science</td>
<td>3.90</td>
<td>0.88</td>
<td>3.48</td>
</tr>
<tr>
<td>Attitudes towards tablets</td>
<td>4.47</td>
<td>0.57</td>
<td>4.51</td>
</tr>
</tbody>
</table>

Note. *p≤.05, **p≤.01. ***p≤.001

**Total immersion: Experienced feelings of presence and flow**

A Mann-Whitney U test indicated statically significant differences in students’ total immersion in both conditions, after the intervention (Table 2). In particular, the students who participated in the strongly-coupled version of the activity outperformed their counterparts, who participated in the loosely-coupled version of the activity, in terms of their reported sense of presence. No differences were detected in terms of flow.

<table>
<thead>
<tr>
<th></th>
<th>Condition 1 (Strongly-coupled AR)</th>
<th>Condition 2 (Loosely-coupled AR)</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Flow</td>
<td>5.65</td>
<td>1.50</td>
<td>5.39</td>
</tr>
<tr>
<td>Presence</td>
<td>5.20</td>
<td>1.57</td>
<td>3.89</td>
</tr>
</tbody>
</table>

Note. *p≤.05, **p≤.01. ***p≤.001
Learning gains: Factual knowledge and reasoning

A Mann-Whitney U test indicated statically significant differences between the students’ learning gains in the two conditions after the intervention (Table 3). In particular, the students who participated in the strongly-coupled version of the educational activity outperformed their counterparts, who participated in the loosely-coupled version of the activity, in their ability to explain their reasoning but not in terms of their factual knowledge.

Table 3: Comparison of learning gains for the two conditions

<table>
<thead>
<tr>
<th>Condition 1 Strongly-coupled AR</th>
<th>Condition 2 Loosely-coupled AR</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Total score</td>
<td>11.77</td>
<td>2.52</td>
</tr>
<tr>
<td>Factual knowledge</td>
<td>7.77</td>
<td>1.45</td>
</tr>
<tr>
<td>Reasoning</td>
<td>4.00</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Note. *p≤.05, **p≤.01. ***p≤.001

Immersive factors: What has contributed to flow and presence?

The factors identified as having affected the students’ sense of flow were classified in two basic themes: (a) Focus, expressing students’ attention during the activity, and (b) Level of challenge, expressing the level of perceived difficulty for the students. The factors identified as affecting students’ sense of presence were classified in two different basic themes: (a) Realism, expressing the extent of the activity’s fidelity for the students, and (b) Agency, expressing students’ perceived sense of control and volitional involvement in the activity. These factors are presented in Table 4 and discussed in the following subsections in relation to the two students’ conditions.

Table 4: Categorization of factors reported as affecting feelings of flow and presence in each condition

<table>
<thead>
<tr>
<th>Organizing themes</th>
<th>Basic themes &amp; Factors discussed</th>
<th>Condition 1</th>
<th>Condition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sense of flow</td>
<td>1.1 Focus</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>1.1.1 Augmentation of reality</td>
<td>33</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1.1.2 Problem-solving process</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.1.3 Narrative plot</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1.4 Navigation and orientation</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1.1.5 Collaboration</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.1.6 Technology hardware</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.1.7 External interruptions</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Level of challenge</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1.2.1 Problem-solving process</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1.2.3 Artifacts and real props</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2. Sense of presence</td>
<td>2.1 Realism</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.1.1 Augmentation of reality</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2.1.2 Video-based characters</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.1.3 Multimedia material</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.1.4 Narrative plot</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.1.5 Artifacts and real props</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.1.6 QR codes arrangement</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>2.2 Agency</td>
<td>14</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2.2.1 Problem-solving process</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.2.2 Narrative plot</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.2.3 Artifacts and real props</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. The (+) and (-) signs indicate whether a factor was positively or negatively evaluated by the students.

Focus

Students in both conditions indicated that their feelings of flow depended on their focus during the activity, and mentioned the following seven factors as attracting or distracting their focus during the learning process: (a) the augmentation of reality, (b) the problem-solving process, (c) the narrative plot, (d) navigation in real space, (e) collaboration, (f) the technological hardware and (g) external interruptions (from other student pairs). The most
frequently reported factor was the problem-solving process, which was equally present and positively evaluated in both conditions. As students mentioned, the problem-solving process increased their focus and attention. E.g.

*I was focused during the activity... As we had to look for evidence, organize them, and then find out the answer. (I was concentrated) as we had to reach a conclusion. From all the available data, we had to reach an outcome. From the first, second, third... From all the data anyway. From all the data you should find a basic... You should combine everything.*

[#13, Girl, Loosely-coupled activity]

Also, in most of the cases students in both conditions emphasized how the narrative plot captured their focus. E.g.

*At the beginning, we felt anxiety. What will we find out?... What is the case about?... And because of the continuous action you feel impatient, to see what is going on until the end.*

[#24, Girl, Strongly-coupled activity]

### Level of challenge

Students in both conditions suggested that their feelings of flow depended on their perceived level of challenge. Students discussed two factors defining the level of difficulty they had experienced: (a) the problem-solving process, (d) the integration of artifacts and real props (only mentioned in the case of the strongly-coupled activity), as in two cases students reported that they had difficulty to understand how the artifacts and real props were related with the case that they had to investigate as well as with the rest of the data that they had already collected. The most positively and negatively evaluated factor in both conditions was the problem-solving process. According to the positive evaluations, the level of the problem-solving was challenging but yet satisfactory. E.g.

*It was difficult but interesting. It was complicated... It was difficult as you had to had to gather many data. During the activity, time flew. It was really nice.*

[#2, Boy, Loosely-coupled activity]

In contrast, according to the negative evaluations, the problem-solving process seemed very easy or very difficult for other students. E.g.

*We were investigating the case.... However, the solution was obvious... It was all about food poisoning. We had no purpose to search anymore for something that was already obvious.*

[#26, Girl, Strongly-coupled activity]

### Realism

Students reported that their feelings of presence depended on the perceived level of realism of the activity. Students in both conditions discussed four factors as positively or negatively contributing to the realism of the activity: (a) the augmentation of reality, (b) the video-based characters, which were activated at each QR code, (c) the multimedia material (e.g., videos, photos, etc.) which were provided during the activity, and (d) the narrative plot. The most positively evaluated factor by students in both conditions was the narrative plot which provided an authentic case for investigation. E.g.

*Generally, it was quite a true story. For instance, it was not about a dragon, which killed the students and then disappeared. It was not very fictional. It was making sense.*

[#16, Boy, Strongly-coupled activity]

A difference between the conditions was that realism was reported as more increased for the students who participated in the strongly-coupled activity due to two additional factors: (a) the artifacts and the real props, as well as (b) the QR codes’ purposeful placement in the physical space. According to the students, the artifacts and real props allowed the collection of “realistic” evidence and made the investigation more “alive”. E.g.

*When we found the notepad, I was carried away as it did not seem as something fake. [...] I thought that it was real, as there were also (handwritten) notes inside.*

[#20, Boy, Strongly-coupled activity]

Students also reported that the QR codes’ purposeful placement in the physical space allowed a more heightened relation between the narrative plot and the real space. E.g.
For instance, when we visited the cafeteria, it was a real cafeteria! It was not something that I would have just seen through the tablet.

[#25, Boy, Strongly-coupled activity]

Both of these factors, which emerged as positive factors only in the case of the strongly-coupled activity, appeared to be also as the most influencing ones for students’ experienced sense of presence.

**Agency**

Students reported that their feelings of presence also depended on their perceived level of agency. Students in both conditions discussed three main factors as positively and negatively contributing to their sense of agency: (a) the problem-solving process, (b) the narrative plot, and (c) the artifacts and real props. According to students in both conditions, the narrative plot positively contributed to their sense of agency as they felt a personal involvement with the case. E.g.

*It was really nice. It felt like that you were “inside”, as you were making the activity. [...] It felt like you were a detective and you were looking to find the truth that was hidden in the mystery.*

[#10, Girl, Loosely-coupled activity]

However, an emerging difference between the conditions was that the sense of agency was more heightened for the students who employed the strongly-coupled activity, due to the artifacts and real props, which allowed students to interact with “realistic” evidence and collect data through a more interactive process. E.g.

*It was different, because in the tablet there were only some images. However, when there were props it was different. You could hold the evidence, you could see it, you could study it better…*

[#24, Girl, Strongly-coupled activity]

In contrast, the lack of artifacts and real props was a negatively evaluated factor for the students who participated in the loosely-coupled activity. Some of these students suggested the integration of real props which could provide more interactivity to the investigation process, even though they were not prompted on this topic. E.g.

*There were no real clues, we were just scanning the QR codes with the tablet. [...] I would prefer the QR codes to be smaller or even hidden to search for them... as well as to provide us with real clues as evidence, and not only the website-based ones.*

[#8, Boy, Loosely-coupled activity]

Overall, the use of artifacts and real props emerged as the most positively influencing factor for the condition with the strongly-coupled version of the AR activity, while their absence emerged as the most negatively influencing factor for the condition with the loosely-coupled version of the activity.

**Discussion and implications**

The present study sought to investigate the hypothesis that the greatest the coupling between the physical space and the narrative of an AR activity is, the greater will students’ immersion and subsequent learning be. In this context, we purposefully compared two conditions; Students in Condition 1 participated in an AR activity with strong coupling between its narrative and the physical space, while students in Condition 2 participated in a loosely-coupled version of the activity.

The findings from the analysis of the data from the two conditions indicated several differences. Students who participated in the strongly-coupled version of the AR activity had no statistically significant difference in their experienced flow when compared to the students who participated in the loosely-coupled version of the activity. However, students in the strongly-coupled version outperformed their counterparts, in terms of the experienced sense of presence. This finding is aligned with the study of Reid et al. (2005), who argued that presence is increased when the virtual world of the activity matches with some artifact of the surrounding physical environment. In contrast, this result contradicts the position of Rossito et al. (2016), who argued in favour of loosely-coupled AR activities.

The analysis of students’ post-activity interviews shed light to our findings in terms of the experienced sense of flow and presence between the two conditions. In particular, the integration of real props as well as the purposeful placement of the QR codes in the physical space did not seem to affect students’ focus as well as their perceived level of challenge; as such, this could explain the lack of statistical difference between the experienced flow among both conditions. On the other hand, the integration of real props as well as the QR codes’ purposeful
placement in the physical space had a significant positive impact on students’ perceived realism, as well as on their perceived sense of agency, which appeared as the most positively influencing factors for presence. As such, this could explain why the students who participated at the strongly-coupled version of the activity outperformed their counterparts, in terms of their experienced feelings of presence. These findings are aligned with prior arguments claiming that place-dependent learning experiences, such as those provided by strongly-coupled educational AR activities, may provide a “narrative” hook as well as an increased sense of agency or sense of control during the activity (Squire et al., 2007). In contrast, a place-independent educational AR activity, even though increasing the portability of the activity, may decrease the amount of authentic interaction with the environment (Klopfer & Sheldon, 2010). Our findings provide empirical support, while they also extend these arguments, as we have found that a strongly-coupled activity can heighten the feelings of presence.

The differences between students’ learning in the two conditions could be attributed to the differences in students’ perceived immersion. According to our findings, the students who employed the strongly-coupled version of the AR activity and experienced higher levels of presence also had higher learning gains, as they were able to explain their reasoning better. These findings provide empirical support for Cheng and Tsai’s (2013) assumption that “perceptions of presence are expected to relate to learners’ behaviors in AR-related learning” (p. 459), thus defining also their subsequent learning gains. Aiming to investigate this argument further, our future work will analyze the learning process of the students who participated in each condition. A better understanding of how learning occurs in informal AR learning contexts can support the creation of hybrid spaces for learning, and the development of more immersive and effective AR learning environments.

References
Refutation Text and Argumentation to Promote Conceptual Change

Christa Asterhan, Hebrew University Jerusalem, Israel, asterhan@huji.ac.il
Maya Resnick, Hebrew University Jerusalem, Israel, MayaResnick@mail.huji.ac.il

Abstract: In the present work, we examine the accumulative effect of two instructional methods for conceptual change, refutation text and argumentation, which are expected to support two complementary processes that, according to current models, underlie conceptual change: Promoting awareness to and reducing interference of irrelevant knowledge structures and sense-making of the counterintuitive, scientifically accepted notions. Hundred undergraduates were randomly assigned to either read a refutation text and then conduct a peer discussion (Ref+Arg), or to read a refutation (RefOnly) or an expository (Control) text, followed by a standard, individual problem solving task. Results showed strong effects for refutation texts on both immediate and delayed post-tests. Contrary to expectations, subsequent peer argumentation did not further improve learning gains. Dyadic dialogue protocols analyses showed that gaining dyads were overall more likely to be symmetrical and to include a discussion of the core principles of natural selection. Several directions for future research are discussed.

Introduction
For more than four decades, scholars from science education, developmental psychology, the learning sciences and cognitive science have documented how children’s and adults’ naïve theories about natural phenomena do not align with the scientifically accepted, but often counter-intuitive concepts that they are exposed to in science instruction. Coming to understand and being able to correctly use these canonical, scientific explanations is not a matter of "gap-filling", in that learners lack the necessary knowledge, but rather involves a substantive re-organization of existing knowledge structures, an outcome which is usually referred to as "conceptual change" (e.g., Chi, 2008; Thagard, 1992; Vosniadou & Brewer, 1994). Current cognitive accounts of conceptual change describe it in terms of a response competition at a deeper cognitive level. Accordingly, it constitutes an increase in the probability with which more advanced schema configurations are activated and used to construct temporary mental representations in working memory, when an individual is required to apply that knowledge to solve a problem (e.g., Potvin, Sauriol, & Riopel, 2015; Ramsburg & Ohissel, 2016; Schnottz & Preuss, 1999). This response competition account is further supported by recent empirical evidence showing that conceptual change involves both an improved capability to construct the correct scientific explanation, as well as more efficient inhibition of automatically activated, but irrelevant schemas and propositions (e.g., Babai, Sekal & Stavy, 2010; Dunbar, Fugelsang & Stein, 2007; Masson, Potvin, Riopel, & Foisy, 2014; Shtulman & Valcarcel, 2012; Potvin, Masson, Lafortune & Cyr, 2015).

In order to be effective, instruction for conceptual change should then preferably support both these cognitive processes: to provide students with opportunities to become aware of and understand the errors in (their) naïve theories, as well to fully comprehend the scientifically accepted theory that is often counterintuitive to everyday experiences (see also Chan, Burtis & Bereiter, 1997). Not surprisingly, traditional tell-and-practice teaching approaches have not been found to be very effective for learning that requires conceptual change, especially in the case of robust misconceptions (Chi, 2008; Vosniadou & Mason, 2013). Researchers of instructional approaches for conceptual change have then studied the effectiveness of alternative instructional techniques, materials and activities, such as refutation texts, argumentation and modeling. Traditionally, each of these have been studied in isolation, however, which may perhaps explain the overall modest effect sizes in this field. In the present work, we examine the accumulative effect of two instructional methods for conceptual change, namely refutation text and argumentation, which are expected to complement each other they support the two aforementioned processes underlying conceptual change: Promoting awareness to and reducing interference of irrelevant knowledge structures (refutation text) and sense-making of the counterintuitive, scientifically accepted notions (collaborative argumentative dialogue).

Refutation texts and conceptual change
Much of what students learn in science classes still comes from textbooks, which are predominantly structured as expository texts of factual information on a scientific concept, without directly referring to common misconceptions (Osborne, 2010; Tippet, 2010). However, research on text reading has shown that expository science texts induce superficial processing and fail to support deep learning (Chambliss, 2002; Diakidoy, Kendeou & Ioannides, 2003). Refutation texts, on the other hand, provide an explicit statement of commonly held
misconceptions, directly refutes them, and then introduces scientific explanations as alternatives (Sinatra & Broughton, 2011; Tippett, 2010; Vosniadou & Mason, 2012). Studies have shown that when learning requires the restructuring of prior incorrect knowledge, refutation texts are generally more beneficial than standard expository science texts (e.g., Braasch, Goldman, & Wiley, 2013; Diakidoy et al., 2003; see Guzzetti et al, 1993 and Tippett, 2010 for reviews). Van den Broek and Kendeou (2008) have proposed that the presentation of the two side-by-side provides the reader with the opportunity to compare the two conceptions, detect inconsistencies and “revise” knowledge structures (the co-activation hypothesis). Findings from eye-tracking data while reading further supports this explanation (e.g., Ariasi & Mason, 2011; Kendeou, 2011).

However, not all studies have found an advantage of refutation over expository texts (see for example, Diakidoy et al., 2016; Hynd & Guzzetti, 1998; Mason, Gava, & Boldrin, 2008; Mason, Zaccheletti, Carretti, Scrimin Palmer, 2003). In addition, the strongest effects for refutation texts are typically reported on less complex topics, such as whether ostriches bury their heads in the sand, or not (see Tippet, 2010; Van Loon et al., 2015). These types of erroneous ideas can be relatively easily refuted by correcting a mistaken belief and do not require revisions or restructuring of complex knowledge systems. Based on the available research, it is not clear to what extent reading a refutation text can in by itself induce conceptual change of robust misconceptions on particularly complex science concepts, as considerably fewer studies have focused on them (but see for example Diakidoy et al., 2003 for an exception). In the present study, we compare the effect of reading a refutation vs. an expository text on a complex science topic for which students are known to have particular robust misconceptions (i.e., natural selection).

Finally, findings from two recent studies (Diakidoy et al., 2016; Van Loon et al., 2015) seem to indicate that refutation texts may be particularly effective in neutralizing the interfering influence of misconceptions, what Van Loon et al. (2015) have termed ‘outdating’. In both studies, findings suggest that refutation texts may not be very effective in helping students to make sense of and construct scientifically correct explanations on their own (‘updating’). Indeed, Vosniadou and Mason (2012) argue that “(r)efutation texts (…) must be used together with other instructional interventions in the context of a rich learning environment that fosters and sustains conceptual change” (p. 35). Accordingly, in order to ensure deep and long-lasting change on complex science topics, students should be given opportunities to make sense of, explain and practice their newly acquired understanding, after reading the refutation text. This proposition is tested in the current study, in which we compare the effects of reading a refutation text on a complex scientific topic (i.e., natural selection) with the effects of refutation text reading followed by subsequent dyadic argumentation.

**Argumentation and conceptual change**

Research has shown that peer argumentation can be an effective means for conceptual change type of learning (see review by Asterhan & Schwarz, 2016). However, productive peer argumentation, when two (or more) discussants compare and weigh different explanations through reasoned argument in a constructive atmosphere (Asterhan & Babichenko, 2015), is not easily elicited nor sustained. Close inspections of dialogue protocols reveal that often times participants simply do not detect when explanations are conceptually different (Sfard, 2009) and therefore may not feel the need to further explore them. The explanations students propose during discussions are often times short, shallow and/or partial. Even when they are using identical terminology, they may implicitly be attaching different meanings to it (Sfard, 2009). This lack of detailed information and ambiguous language use may create an ‘illusion of consensus’ between dialogue participants, believing that they are in fact proposing similar solutions and are in agreement.

Task design should then make sure that the differences between misconceptions and scientifically accepted theories are presented in a more salient and comprehensible manner so that student have access to them during argumentation. In the present study, we support student argumentation on scientific concepts by presenting them with erroneous worked-out examples (Asterhan et al., 2015; Durkin & Rittle-Johnsson, 2011) that are based on common misconceptions. Student dyads are asked to solve and correct through argumentation, which is expected to further help them in inhibiting irrelevant knowledge structures, as well as with understanding, consolidating, and strengthening the correct concept (‘updating’).

**Hypotheses**

Based on the aforementioned rationale the following hypotheses are formulated:

H1: Students who read a refutation text on natural selection will show larger learning gains than students who read an expository text on the same topic.
H2: Students who read a refutation text and then participate in peer argumentation on erroneous worked-out examples will show greater conceptual gains than students who will read refutation texts and then complete a standard individual problem-solving activity.

H3: Reading refutation texts will improve students’ capability to identify and correct errors (‘outdating’), whereas peer argumentation will further improve students’ capability to identify and construct correct explanations (‘updating’).

H4: Dyads who conduct a critical discussion on the core principles of natural selection will show greater gains from the peer collaboration session than dyads who do not.

**Method**

**Participants**

One hundred undergraduate students (73 F, Mage = 24.5) from a large university in Israel participated in the study. Requirements for participation were a lack of formal background in the Life and Exact Sciences and proficiency in the Hebrew language. Participants could choose between receiving course credit (31%) or financial reimbursement (approximately $15) for participation. Four participants failed to appear for the delayed posttest (2 from the Ref+Arg and 2 from the RefOnly condition). The pretest and immediate posttest scores of these participants were not found to be different from the remainder and their data was therefore omitted only from analyses that include the delayed posttest scores.

**Design**

A 1X3 between-subject experimental design was used. Participants were randomly assigned to one of three conditions (see Figure 1): (1) Refutation text + dyadic argumentation on erroneous solutions (Ref+Arg; N = 50); (2) Refutation text + individual problem solving (RefOnly; N = 26); (3) Expository text + individual problem solving (Control; N = 24). Individual conceptual understanding of natural selection was assessed on pre-test, immediate post-test (following text reading), and delayed post-test (a week later).

**Tools**

*Demographics and background.* The background questionnaire targeted the following demographic information: gender, age, degree, field of major, religious affiliation, degree of religiosity, and background in high school biology education and evolution. In addition, students' attitudes and beliefs regarding the theory of natural selection were assessed with 4 Likert scale items, ranging from 1 ("I do not agree at all") to 5 ("I completely agree"), which were translated to Hebrew from Shulman (2006). Examples are "Natural selection is the best explanation for the creation of species" and "Species change over time". Internal reliability for the attitude measure was high, Cronbach's α = .84.

*Conceptual understanding of natural selection.* Individual conceptual understanding of natural selection at pre-test, immediate post-test and delayed post-test was assessed with open and closed items that targeted the evolution of selected animal traits. Test items were adapted from Asterhan et al. (2015). Eight different animals and traits were distributed over the three tests. The pretest and the delayed posttest each included questions about 3 different animals: Six different false/correct statements about the first animal, five about a second animal and one open-ended question about the third. The immediate post-test included questions about 2 animals: Six true/false statements about the first animal and one open-ended question about a second animal. For each animal, students were presented with a short text about a specific animal species, a physiological change in a specific trait over time and a short description about the importance of that trait. In the false/correct items, this text was followed by five (or six) statements, which each addressed a different principle of natural selection (e.g., intra-species variability, proportional change). Students were required to indicate whether the statement was false or correct and then explain and justify their choice in their own words. In the pretest and delayed posttest, which included sets of false/correct statements on 2 different animals, when a given principle was presented as a correct statement for the first animal, the principle was presented as an incorrect statement for the second animal. For a given animal, approximately half of the statements presented were correct and half incorrect. In the open item questions, students
were required to give a full explanation of how they think the given trait had evolved, according to the theory of natural selection.

**Instructional texts.** Two instructional texts on the topic of natural selection were created: an expository text (453 words in Hebrew) and a refutation text (525 words in Hebrew). Both were identical with regard to the background information and the correct explanations of natural selection, which related to the existence of differences within species and to the evolution of traits over time. The relevant terminology was explained in plain language and they included an elaborated description of a well-known example of change in a specific feature of a specific species, namely the increasing length of the giraffe's neck. The only difference between the texts was that while the expository text only presented the scientifically accepted theory, explanation and example, the refutation text in addition included (1) explicit references to common misconceptions (intentional change, intra-species diversity and the source of diversity and change); and (2) explicit statements that those are "wrong" (the refutation cue). The remainder of the text was verbatim identical.

**Worksheets.** The worksheet booklet students used in stage 4 of the experiment was based on materials used in a previous study (Asterhan et al., 2015), and consisted of three open-ended questions about two novel evolutionary phenomena, namely the webbed feet of ducks and the wing coloring of the peppered moth. Each question appeared on a separate page. The first two questions were textual and similar in format to the conceptual knowledge test open items, which required students to explain the described change in terms of natural selection. In the third task, which was adapted from Shulmann (2006), students were asked to depict the gradual change in wing coloring of the peppered moth population in a graphical manner. This item is used to distinguish between typological and selection-based representation of change.

Students in the Ref+Arg condition received the same booklet, with two changes: First, the space for writing the solutions was already filled with a solution provided by a (fictitious) peer student. The textual solutions were handwritten (each in a different handwriting) or hand-coloured. They were erroneous, targeting a particular common misconception in each of the three solutions and adapted from common student answers from previous data bases. The errors in the two textual solutions were highlighted with a yellow marker. For example, the solution to the webbed duck feet question was designed to refer to the common misconceptions that individual animals intentionally changed a trait during their lifetime and that acquired changes in traits are passed on (highlighted error in italics here):

"The ducks needed webs to swim. They had to know how to swim in order to survive. Some of the ducks managed to develop webbed feet for themselves. They survived and managed to reproduce. Those ducks that did not manage to develop webbed feet- did not survive"

The erroneous textual solution to the moths question alluded to misconceptions about existing intra-species variability ("before all the moths were white") and a typological change ("in each generation every moth became a bit darker"). Finally, the graphic depiction was already coloured by a (fictitious) peer, to depict a classic typological model of change. Second, a separate space was reserved at the bottom of each page for the participants to fill in the corrected solution in their own handwriting and their own colouring.

**Procedure**

Except for stage 4 (problem solving stage), all stages of the experimental sessions were conducted in individual, separate rooms. Following a brief verbal explanation about the experiment, students filled in the background information survey, followed by the pretest. They were instructed to answer in full and elaborate their answers as much as possible, even when they were uncertain of their answers. Following (stage 2), each participant received either the refutation text (Ref+Arg and RefOnly conditions) or the expository text (control condition) about natural selection. They were told that the text presents the scientifically accepted explanation and received 10 min to study the text in detail. Following the reading phase, participants completed the immediate posttest individually (stage 3). They were then moved to a different room shared with another participant (stage 4). Assignment to peer participant was random within condition. Participants in the RefOnly and the control condition were told that the reasons for this move were logistic as the room was needed for another experiment. They were seated with their backs to one another and received the standard worksheet with open questions, which they were instructed to solve individually and without talking to each other. Participants in the Ref+Arg condition were instructed to work in pairs to critically, yet constructively, discuss the erroneous student solutions in the filled-in worksheets. The dyadic argumentation sessions were audio-recorded. Upon completion of the discussion, each participant received a copy of the worksheets they previously discussed with room to correct the presented mistakes individually. A week later, participants returned for the final stage to complete the delayed posttest questionnaire and received a debriefing and reimbursement for their participation.
Coding

Coding understanding of evolution. Coding of students’ written solutions to the test items was based on existing coding procedures developed in previous studies (Asterhan et al., 2015; Asterhan & Dotan, in press). Each written solution was graded according to accuracy and compliance with the main principles of natural selection: 0 for omissions, misconceptions or other crucial errors, .5 for partially correct solutions, or full credit (1) for solutions that contained no misconceptions and addressed the main tenets of natural selection correctly.

Each false/correct item targeted one of the six predefined principles of natural selection (Intra-species diversity, Source of diversity and change, Inheritance of traits, Learnt behaviours, Survival, Proportional vs typological change). When coding for the closed questions, two factors were considered in one single score: the indicated choice of right or wrong and the accompanying textual explanation. A correct choice between right and wrong together with a correct and sufficient explanation resulted in full credit. An incorrect right/wrong choice with an incorrect explanation, resulted in zero points. When these two components were not synchronized, more weight was given to the verbal explanation. Most of these cases indeed showed partial understanding and were given a score of 0.5, but there were several cases that showed clear misconceptions in the written explanations in spite of choosing the correct answer, and received 0 points.

Written solutions to the open items were coded in a similar manner, but regarded the overall model of evolutionary change represented in the student answers (see Asterhan & Dotan, in press), instead of a particular principle. Solutions that contained no misconceptions and correctly explained change in terms of existing variability, selection and proportional change received full credit (1). Answers that were partially correct or presented both correct as well as incorrect aspects received .5 points. This is also the case for well-documented hybrid models (Asterhan & Schwarz, 2007), such as solutions that refer to existing intra-species variability on an “ability to change” together with a selection mechanism of those members who had managed to change themselves. After a period of training and discussion, three human coders scored the same 248 randomly chosen item responses. Interrater reliability was good, .72 < Cohen’s Kappa < .79. Differences were resolved through discussion, after which the entire data set was coded. A total conceptual understanding score was compiled by adding the different scores for each test item on each test, while assigning the open test item score a weight of 5 points (instead of 1).

Coding of dialogues. Discussions of the 25 dyads in the Ref+Arg condition were audio recorded. Twenty-three discussions could be transcribed (1 was incomprehensible and 1 was never recorded due to technical failure). The mean length of these audio recorded discussions was 8 min and 16 sec. Transcriptions included all verbal content, as well as significantly long pauses, and laughing, but not intonation and other auditory features. Discussion protocols were coded as a whole. Following findings from previous work on the association between dyadic argumentation and conceptual change (e.g., Asterhan & Schwarz, 2009), initial coding efforts focused on two discussion characteristics: whether the discussion could be characterized as critical, dialectical overall and whether the interaction was symmetric. A third characteristic, rhetoric style (disputative or deliberative argumentation, Asterhan & Babichenko, 2015), proved to be irrelevant. Perhaps due to the fact that the argumentation instructions were explicitly modelled on and directed toward deliberation, clear cases of disputative argumentation were near non-existent.

This procedure yielded and additional coding category, namely the extent to which student dyads discussed the (six) core conceptual principles of natural selection or not.

In sum, three dialogue characteristics were coded: (1) Critical discourse (0, 1) – when the students overtly confronted the different solutions and related to the differences by providing justifications, explanations and counterargument, it was considered a critical discussion; (2) Symmetry (0,1) – when the word count from all the conversational turns from each of the two discussants exceeds 35 % of the total discussion word count (excl. repetitions) the discussion was deemed symmetrical; and (3) Discussion of core principles (0, 1) – when at least 5 of the 6 principles of natural selection came up during the discussion the grade 1 was assigned, when 4 or less core principles were mentioned it received the grade 0 (N. B. the erroneous examples referred to three core principles altogether). Two raters scored all the dialogue protocols independently, Cohen’s Kappa = .75.

Results

Normalized (Hake) gain scores were computed for the overall conceptual understanding score, as well as for the false statement items score and for the correct statement items score separately. No differences were found on pretest scores between the different conditions, $F(2, 97) = .60, p = .560$. Neither of the three control variables (i.e., attitudes toward natural selection, religiosity, and perceived understanding of evolution) yielded differences across conditions, nor did they correlate with normalized pre-to delayed posttest gain scores ($r = .13, r = -.17$ and $r = -.09, ns$, respectively).
Effects of condition on conceptual understanding

All univariate analyses of variance presented here were conducted on the normalized gain scores of the full data set. Levene's tests for equality of variance across compared conditions showed that this assumption was not violated in any of the reported statistical analyses ($p > .80$). To test the stability of the results and the reliability of the chosen method for analyses, we reran each comparison with two additional statistical models (ANOVA on raw delta scores and ANCOVA with pretest or immediate posttest as covariates), as well as on the data set while excluding (eight) participants with high pretest scores ($r > .85$; two in the control, two in the RefOnly and four in the Ref+Arg condition). These yielded identical results and are therefore not reported on further.

The effect of refutation text on conceptual gains. In order to examine the effect of text type (refutation vs. expository) on conceptual knowledge, mean normalized gain scores from pretest to immediate posttest were compared between students who had read a refutation vs. an expository text. Students in the refutation text condition showed larger conceptual gains on the immediate posttest ($M = 43.21, SD = 44.86, N = 74$) than students in the expository text condition ($M = -1.94, SD = 52.77, N = 26$), $F (1, 98) = 17.75, p < .001$, with a large effect size of $\eta^2_p = .15$. Further analyses showed that this advantage was also evident on the delayed posttest: Students who had read a refutation text showed larger conceptual gains ($M = 46.10, SD = 36.53, N = 70$) than students who had read an expository text ($M = -1.94, SD = 52.77, N = 26$), $F (1, 94) = 18.10, p < .001$, with a large effect size of $\eta^2_p = .16$. These findings corroborate the first hypothesis (H1), according to which refutation texts are more effective than expository texts for both short-term and long-term conceptual gains.

The additional effect of argumentation on conceptual understanding. A one-way ANOVA compared the mean normalized gain scores from pretest to delayed posttest, across the three conditions. A main effect of condition on normalized gains was found, $F (2, 93) = 9.02, p < .001$, with a large effect size of $\eta^2_p = .16$. Post-hoc tests with Tukey corrections showed that the gains in the control condition ($M = 9.32, SD = 40.60$) was significantly lower than gains in the RefOnly condition ($M = 48.37, SD = 35.55, p < .002$) and gains in the Ref+Arg condition ($M = 45.06, SD = 37.29, p < .001$). However, no differences on normalized gain scores were found between the latter two, $p = .938$. A one-way ANOVA compared the mean normalized gain scores from immediate to delayed posttest, across the three conditions. Overall, gains from the additional activity were low with a large variance ($M = 10.26, SD = 34.71$) and no significant differences were found between the normalized gain scores of the Ref+Arg ($M = 14.40, SD = 35.16$) the RefOnly ($M = 8.71, SD = 38.60$), and the control condition ($M = 3.94, SD = 30.41$), $F < 1.01$.

Taken together, these findings do not support the second hypothesis, according to which peer argumentation on erroneous solutions would further increase learning gains compared to individual problem solving activities and result in a significant difference between the RefOnly and the Ref+Arg conditions. Even though immediate posttest scores in these two conditions were overall fairly high ($M = 65.63$ and $M = 73.14$, respectively), there was definitely room for more improvement. It therefore does not seem that the lack of effect could be attributed to a ceiling effect.

### Table 1. Mean normalized gain scores (and SD) for conceptual understanding, per experimental condition

<table>
<thead>
<tr>
<th>Normalized gain score</th>
<th>Ref+Arg</th>
<th>RefOnly</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test → immediate post-test</td>
<td>39.86 (48.62)</td>
<td>50.18 (35.73)</td>
<td>-1.94 (52.77)</td>
</tr>
<tr>
<td>Immediate → delayed post-test</td>
<td>14.40 (35.16)</td>
<td>8.71 (38.60)</td>
<td>3.94 (30.41)</td>
</tr>
<tr>
<td>Pre-test → delayed post-test</td>
<td>45.06 (37.29)</td>
<td>48.37 (35.55)</td>
<td>9.32 (40.60)</td>
</tr>
</tbody>
</table>

Outdating and updating of conceptual knowledge.

In order to test the hypotheses regarding effects of condition on out-dating of misconceived knowledge and on up-dating of correct, scientific knowledge (H3), the following outcome variables were considered: Improvement on the mean normalized gain score of the false statement items only was considered as an indication of outdating, whereas improvement on the mean normalized gain score of the true statement items only was considered as an indication updating and (van Loon et al, 2015). In addition, improvement on the open question item was also considered as an indication of updating (Kendeou & van den Broek, 2005).

Performance on true vs. false statement items. The normalized gain score from pretest to delayed test was calculated separately for the true statement items only and for the false statement items only. As the immediate posttest included only 6 (instead of 11) forced choice questions, analyses were restricted to the pre-test and the delayed post-test scores on this measure. A main effect of condition was found on the normalized false statement gain score, $F (2, 93) = 8.36, p < .001$, with a large effect size of $\eta^2_p = .15$. In alignment with hypothesis 3a, learners who had read an expository text showed significantly less improvement on the recognition and correction of false statements ($M = 17.49, SD = 30.03$), when compared to both the RefOnly condition ($M = 51.21, SD = 29.76, p =$
.002) and to the Ref+Arg condition (M = 17.49, SD = 30.03, p = .001). No differences were found between the two refutation text conditions, p = .922.

A main effect was also found on the normalized true statement gain score, F(2, 93) = 7.27, p = .001, with a large effect size of ηp = .14. The pattern of differences across conditions mirrored the false statement gain score: Learners who had read a refutation text improved significantly more on true statement test items (M = 47.38, SD = 51.29 for the Ref+Arg condition and M = 56.21, SD = 45.00 for the RefOnly condition) than learners in the control condition (M = 6.02, SD = 55.70), p = .004 and p = .003, respectively. Contrary to expectations, however, the addition of a dyadic argumentation activity did not result in larger gains, as no differences were found between the gains in the Ref+Arg and the RefOnly condition, p = .781.

Open Question Items. Evidence of substantive conceptual gains (conceptual change) in student explanations to the open-ended item was defined as an improvement of .5 on the nominal, unweighted open item test score (see Asterhan & Dotan, in press). Only students who had a nominal, unweighted pretest open item score of .5 or lower were included in the analyses (N = 73). Seventy-five per cent of students in the refutation text conditions showed substantive conceptual gains from pretest to delayed posttest, whereas only 45% of students in the control condition did. A Chi square test showed this difference to be significant, χ² (1, 69) = 5.84, p = .016, suggesting that the experimental intervention also resulted in greater improvement on the open items. A Chi square test for the differences in conceptual gains from pretest to immediate posttest, following only the reading intervention, also showed a strong significant difference between refutation text (75% improved) and expository text readers (27% improved), χ² (1, 72) = 14.32, p < .001. A comparison between the Ref+Arg and RefOnly conditions only showed no significant differences in improvement from pretest to delayed posttest, χ² (1, 47) = 1.21, p = .271, nor from immediate to delayed posttest, χ² (1, 30) = 2.44, p = .118.

Taken together, this set of analyses suggests that refutation texts improved not only outdating, but also updating processes, which was evident on both immediate as well as delayed tests. Argumentation with a peer did not further improve students’ capability of recognizing and formulating correct scientific explanations of natural selection processes further, compared to an additional individual problem solving activity.

Dialogue protocol analyses
In order to obtain further insight in the discussion features that were associated with learning gains, the 23 available dyadic dialogue protocols in the Ref+Arg condition were analysed according to the following three criteria: Contribution symmetricity, critical discourse, and reference to core conceptual principles (see Coding section). The discussion characteristics of dyads in which none of the dyad partners showed a substantive gain from the immediate to delayed post-test were compared to dyadic discussions in which at least one dyad partners showed such gains. In three dyads, both partners had near perfect scores (> 91) and these were not included in the discussion feature analyses. Substantive gains were defined as a further increase of 30% from the immediate to the delayed post-test (i.e., normalized gain score > 30). This definition resulted in 10 “gaining” and 10 “nongaining” dyads. Table 2 presents the cross-tabulation of dialogue features and dyad gains. Dialogues of gaining dyads were more likely to include references to core conceptual principles of natural selection, χ² (1, 20) = 7.20, p = .007. Their interactions also tended to be symmetrical more often, even though this difference was only marginally significant, χ² (1, 20) = 3.33, p = .068. A certain trend could be observed by which gaining dyads’ discussions seemed to be more often characterized as critical (8 out of 10), but this was not significant, χ² (1, 20) = 1.98, p = .160. Even though 8 out of 10 gaining dyads conducted a critical discussion, 5 of these did not result in further gains.

Table 2. Dialogue features of gaining and non-gaining dyads

<table>
<thead>
<tr>
<th></th>
<th>Neither partner gains (N = 10)</th>
<th>At least one partner gains (N = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical discussion</td>
<td>Yes 50%</td>
<td>80%</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Yes 20%</td>
<td>60%</td>
</tr>
<tr>
<td>Nr. of core principles</td>
<td>&gt;5 20%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Discussion
The results presented here show, first and foremost, that refutation texts are an effective means to improve conceptual understanding, not only in the case of simple, unidimensional beliefs, but also for multi-dimensional, complex science topics, such as natural selection. This advantage was evidenced immediately following the text reading, but also on delayed post-tests, a week later, suggesting that its effect was not superficial, but caused a
substantive improvement. Moreover, and in contrast to previous findings (Diakidoy et al., 2016; van Loon et al., 2015) we found that refutation texts not only improved students’ capability of detecting and correcting misconceived ideas (outdating), but also their capability to construct and identify correct explanations (updating).

The expectation that subsequent sense-making activities, in the form of dyadic argumentation on erroneous examples, would further improve conceptual understanding over and above the effect of refutation text was not confirmed by the findings. As the mean scores on the immediate test did not exceed 75%, the lack of effect cannot be attributed to a ceiling effect. Further analyses of the dyadic dialogue protocols revealed that in ten of the total of 20 dyads in which at least one student could improve substantially, this in fact happened. Their dialogues were overall more likely to be symmetrical and include a discussion of the core principles of natural selection than the dialogues of those in which neither partner showed substantive gains.

Limitations and future directions
Future research should extend this research to a population of younger, school-aged children and outside of university settings in order to investigate whether refutation text without further collaborative sense-making processes could be expected to be equally effective in these age groups. It is likely that the effects of refutation texts would be weaker and that the added sense-making activities would have a larger impact in those age groups. Second, the present study did not include a condition in which argumentation preceded refutation text reading. It is therefore not possible to draw any conclusions concerning the relative effectiveness of either instructional activity in isolation (i.e., whether refutation texts are more effective than argumentation). This could be explored in future studies.

Finally, it could be argued that the strong, positive effects of refutation texts in this study could be attributed to the testing format, as the tests included a large number of true/false statements, which were presented on the same test page. This is in some ways similar to the refutation text, in which correct and misconceived knowledge structures are directly compared. Findings from previous research have also suggested that refutation texts promote performance on closed test items, but not on open items (Kendeou & van den Broek, 2005). However, in the present study student were required to elaborate their choice and these elaborations were used as the main source for grading (as opposed to the actual false/true choice). Moreover, we also included open test items in our assessment tools. Findings from separate analyses of the open test performance mirrored those of the other measures: Refutation text reading resulted in substantively more improvement than expository text, whereas subsequent argumentation did not further add to these gains. We are therefore fairly confident that the effect of refutation texts on conceptual understanding should not be attributed to test format.

Selected references
Teaching For Versus Through Problem Solving:
Impact on Teaching and Learning

Annelies Raes, Celine Vens, Pieter Vanneste, and Fien Depaepe
Annelies.raes@kuleuven.be, Celine.vens@kuleuven.be, Pieter.vanneste@kuleuven.be, Fien.depaepe@kuleuven.be
ITEC, KULeuven, Belgium

Abstract: This paper presents the results of a quasi-experimental study with third bachelor students in Medicine comparing the teaching for problem solving and teaching through problem solving condition. This study focuses both on the outcome and process effects. On the cognitive level this includes students’ conceptual learning gain. On the affective level this includes students’ experienced motivation both in the instruction phase and the problem-solving phase. This study complements previous research by taking a zoomed-in perspective including multimodal data streams (i.e. audiovisual, self-report, logdata and psycho-physiological data) to fully understand the effects on the teaching and learning mechanisms. Although we did not find statistically significant effects regarding the conceptual understanding, the results of the affective outcome and process variables are consistent with the growing body of evidence that generating solutions to novel problems prior to instruction can improve deeper learning and can make teaching more interactive and student-centered.

Theoretical and empirical framework
Problem solving is often suggested as the new way of learning to cognitively activate students and prepare them for the 21st Century skills (Felmer, Pehkonen, Kilpatrick, 2016; Lester & Cai, 2016). During the past 30 years there have been significant advances in our understanding of the cognitive, metacognitive, and affective aspects of problem solving in domains as science and mathematics. It is known that problem solving has many learning advantages, but it is also known that problem solving places higher cognitive demands and is often difficult for novices in a certain field with inconsistent and limited prior knowledge as indicated within the instructionist-constructivist debate (Kirschner, Sweller, & Clark, 2006). To reduce this burden on the cognitive resources and mitigate problems associated with learner disengagement and learner frustration, many teachers still opt for direct instruction or knowledge transfer preventing struggle and saving time, and to make sure the correct information is taught (Sweller & Chandler, 1991). Although there are several reasons to believe in the effectiveness of problem solving based on considerable research on teaching problem solving in classrooms, there remain far more questions than there are consistent answers about this complex form of activity. One of these questions is if problem solving should be taught as a separate topic in the curriculum or if it should be integrated throughout the curriculum? Schroeder and Lester (1989) identified these two opposite types of approaches to problem solving as 1) ‘teaching for problem solving’ (further abbreviated as TforPS) starting with the content and concepts and then moving to solving problems and 2) ‘teaching through problem solving’ (abbreviated as TthroughPS) meaning students learn the content and concepts through real contexts, problems, situations and problems followed by an instruction phase. Although not equivalent, the latter approach can in a way be related to the theoretical framework of Kapur (2008) defining productive failure as a teaching and learning approach which engages students in solving problems requiring concepts they have yet to learn, followed by a consolidation and instruction phase on the targeted concepts. This approach is based on the work of Schmidt and Bjork (1992) suggesting that experimental manipulations that may hinder performance in the shorter term can actually be productive for learning in the longer term. They even concluded that maximizing performance in the initial learning may not necessarily be the ones that maximize learning in the longer term. Previous research has found that starting with problem solving prompts the activation of prior knowledge and idea generation from different perspectives (Schwartz, Chase, Oppezzo, & Chin, 2011). These studies moreover indicated that approaches as productive failure are specifically beneficial for retention and transfer (see e.g. Kapur, 2013). In recent work, Kapur (2016) deepens this approach and suggests to distinguish four possibilities for designing learning: i.e. productive or unproductive success, and productive or unproductive failure. Within our conceptualization, we assume that the TforPS has a higher chance to lead to productive success, whereas TthroughPS has a higher chance to lead to productive failure.

This study compared the TforPS and TthroughPS condition in a quasi-experimental study with third bachelor students in medicine. This study complements to previous work as it aims to better understand the effects of the opposite approaches on both outcome and process variables including multiple streams of data.
Not only the student’s knowledge gain and behavior is being studied, but also their psycho-physiological state is being monitored. During the experiment, wearables attached to the teacher’s and some of the students’ wrists, measure their skin conductance level, skin conductance phasic responses and skin temperature. In doing so, the relative stress level a student experiences during a certain moment in the experiment can be determined.

Research questions and hypotheses

The first research question focuses on the effect of both approaches on the outcome variables on the cognitive and on the affective level. The outcome variables include students’ conceptual knowledge gain and their self-efficacy on the cognitive level and students’ experienced mental effort, experienced competence satisfaction, and intrinsic motivation and perceived value of the learning activities (instruction phase and problem-solving phase) on the affective level. Thanks to the wearable data, also the stress level that students and the teacher experience is studied. Based on previous research, the hypothesis of this study is that in the TthroughPS condition, on the one hand, the cognitive demands will be higher both for students and for the teacher, but on the other hand, students’ agency and therefore also their interest, during the task and during the following instruction will be higher. Regarding the effect on the learning gain, we did not expect significant differences on the posttest immediately after the intervention, but hypothesize based on previous literature that the benefits of TthroughPS will unfold on the retention test on the longer term (Schwartz et al., 2011). As the effects are highly dependent on the way the intervention is implemented and the way it includes the expected cognitive mechanisms that facilitate learning from productive failure (Loibl, Roll, & Rummel, 2016), evaluating these mechanism was part of our second research question.

The second research question focuses on the effect of both approaches on the process variables and aims to more qualitatively analyze the effects on the learning mechanisms during the problem-solving phase and the teaching mechanisms during the instruction phase. Within the scope of this study and based on the theoretical assumption that TthroughPS is more time consuming and more demanding, we compared the duration and effort during the problem-solving phase in both conditions and hypothesized that starting with the problem solving phase is more engaging, but also more demanding, and consequently more time consuming. Based on previous literature indicating that instruction after problem solving provides more opportunities for interactions between students and teacher, both the amount and duration of the interactions in the instruction phases of both conditions were analyzed and compared.

Methodology: Intervention and measurements

This study was conducted during students’ regular class periods of the course ‘Biostatistics’ taught at the University of Leuven, campus Kulak Kortrijk in Belgium. The specific study was scheduled September 29, 2017. During previous academic years, the content about sampling methods and bias within sampling was given during a lecture based on a power point presentation in which the content is taught using a case study of the National Health Study. Within the scope of the current study, this specific lecture has been redesigned and a transfer task has been added to the instruction phase in which the students were asked to set up a survey to evaluate the psychosocial wellbeing of students within their campus. The focus of the task was not to formulate the questions of the survey, but students had to focus on how to select the method of data collection, how to select a good sample, how to deal with randomness, how to deal with non-response and how to solve the problem of over- and under-sampling through weights. This year, this curriculum was given to 48 third year bachelor students in Medicine. For the purpose of the study, as depicted in Figure 1, students were randomly assigned to one of the two conditions: TforPS and TthroughPS.
In the TforPS condition, students first got direct instruction in which the instructor modeled and contrasted the different sampling methods with appropriate instructional facilitation and explanation to make them attend to critical features and possible bias within the sampling methods. The direct instruction phase was followed by the collaborative problem-solving phase. In the TthroughPS condition the instruction phase followed the problem-solving phase and aimed to explain the exact same content but afforded opportunities to compare and contrast relevant student-generated solutions. The learning setting of this study is illustrated in Figure 2.

In this study the main dependent outcome variables were situated on the cognitive and affective level. To assess students’ knowledge gain, students administered three equivalent test forms, one as pretest, one as posttest, and one as retention test two months after the intervention. In line with previous research (e.g. Kapur, 2013), the pretest included 4 multiple-choice questions and one open-response question; the post- and retention test included 6 multiple-choice questions and the same open-response question. Based on Bandura’s work (1997), students’ perceived self-efficacy was also measured by asking students to rate on a scale from 0 – 100 how confident they felt about the correctness of their given answer. To measure students’ mental effort and motivation during both the instruction and problem-solving phase the Intrinsic Motivation Inventory (IMI) (Deci & Ryan, 2000) was administered immediately after the given session.

To assess and compare the learning and teaching mechanisms, in both conditions, different data streams have been triangulated. First, audiovisual data have been recorded during the complete intervention from three camera positions to capture student-student and student-teacher interactions for observing and capturing the teacher’s and students’ actions and reactions throughout the instruction phase. One camera moreover captured the student-student interactions of one group in both conditions. Second, students’ reasoning and answers during the problem solving phase were logged on the technology platform WISE on which the problem solving task was provided to students and students’ answers were evaluated by means of the knowledge integration rubric (Slotta & Linn, 2009). Third, psycho-physiological data were collected by using individual sensors measuring skin conductance, skin temperature and acceleration.

This overall study design has been evaluated and approved by the Social and Societal Ethics Committee and all students gave their informed consent.
Results

RQ 1. Effects on students’ progress in conceptual understanding, self-efficacy and motivation

Based on paired t-tests, an overall increase between pretest and posttest was found with respect to students’ conceptual knowledge ($t(46)=2.14, p < .001$) and regarding students’ self-efficacy ($t(27)=11.56, p < .001$). Moreover, one-way analyses of covariance (ANCOVA’s) were conducted with posttest scores as dependent variable, condition as independent factor, and pretest scores as covariate to discover whether there are differences between both conditions on the posttest measure regarding conceptual knowledge and self-efficacy, after adjustment for the pretest scores. ANCOVA indicated no significant difference between the conditions as depicted in Figure 3. Moreover, no effect was found on the retention test that was administered two months later.
To test students’ motivation during both the instruction and problem-solving phase the IMI was administered immediately after the given session. Figure 4 displays the results regarding the different subscales of the IMI, including the experienced enjoyment during respectively the instruction and the problem-solving phase, the perceived competence, the effort they put into the activity, the perceived value or importance of the instruction and problem solving, the perceived autonomy or choice and finally the relatedness students experienced. First, we found significant within-subject differences in motivation between the evaluation of the instruction phase and the evaluation of the problem-solving phase. The mean interest and relatedness for example was significantly higher during the group work (respectively ($t(46)=0.29$, $p=0.045$ and $t(46)=0.55$, $p<0.001$). Yet, the perceived value and competence was significantly higher regarding the instruction phase. Although these are interesting results, the main focus of this study was the ordering effect the problem-solving task first versus second. Based on between-subject analyses we found two trends. Regarding the instruction phase on the one hand, we found that students’ perceived choice or experienced autonomy was higher during the instruction given after the group work. On the other hand, we found a lower perceived competence during the problem solving phase given before the instruction phase. It is important to interpret these results with caution as they are based on a small scale study. Nevertheless, these trends are in line with previous findings in the literature.

RQ 2. Effects on the learning and teaching mechanisms

To assess and compare the learning and teaching mechanisms in both conditions, first audiovisual data have been analyzed. Table 1 summarizes the results of the analysis and indicates that in both conditions the time spent to the instruction phase was comparable (i.e. circa 1 hour). However, the time students spent on the problem-solving task differed significantly between both conditions. In the TforPS condition in which the
instruction was followed by the group work, one group already finished the task after 25 minutes. The maximum time spent on the group work was 41 minutes. The time spent on the problem-solving task in the TThroughPS condition was higher (mean time spent was 46 minutes versus 36 minutes in the TForPS condition) and the variation was lower. The minimum time spent on the task was 44 minutes.

When having a closer look at the quality of answers during the problem-solving task, we found a substantial difference between both conditions. To indicate this difference, Table 1 displays the differences in lengths of the answer on the most challenging subtask of the group work. This subtask asked to postulate their sampling method to select 120 students for the campus survey keeping in mind the prerequisite of selecting the same number of students from all students years and guaranteeing the presence of students from large as well as small faculties in the sample.

Table 1: Comparing the duration and interactions in both conditions

<table>
<thead>
<tr>
<th></th>
<th>Teaching FOR PS Instruction followed by group work: N = 23; 5 groups</th>
<th>Teaching THROUGH PS Group work followed by instruction. N = 25; 5 groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of instruction phase</td>
<td>59 minutes</td>
<td>1h4minutes</td>
</tr>
<tr>
<td>Duration of problem solving phase of five groups</td>
<td>Min. time spent:25 minutes</td>
<td>Min. time spent: 44 minutes</td>
</tr>
<tr>
<td></td>
<td>Max. time spent:41 minutes</td>
<td>Max. time spent: 47 minutes</td>
</tr>
<tr>
<td></td>
<td>Mean time spent:36.6 minutes</td>
<td>Mean time spent: 46 minutes</td>
</tr>
<tr>
<td>Student-teacher interactions</td>
<td>3 interactions</td>
<td>6 interactions</td>
</tr>
<tr>
<td></td>
<td>2 initiated by the teacher</td>
<td>5 initiates by the teacher, 1 initiated by student asking a question</td>
</tr>
<tr>
<td></td>
<td>1 initiated by a student asking for a clarification</td>
<td></td>
</tr>
<tr>
<td>Student-teacher interaction time</td>
<td>2min10sec</td>
<td>10min20sec</td>
</tr>
<tr>
<td>Number of students interacting with teacher</td>
<td>2 out of 23 students</td>
<td>6 out of 25 students</td>
</tr>
<tr>
<td>Quality of group work: indicator length of reasoning regarding task 3.1 ‘Sampling’)</td>
<td>Min. 9 words</td>
<td>Min: 43 words</td>
</tr>
<tr>
<td></td>
<td>Max. 44 words</td>
<td>Max: 171 words</td>
</tr>
<tr>
<td></td>
<td>Mean: 24</td>
<td>Mean: 78 words</td>
</tr>
</tbody>
</table>

First, it was revealed that the mean length of answers of student groups within the TForPS condition was significantly lower than the lengths of answers of student groups within the TThroughPS condition ($M=24$ versus $M=78$). Within the TForPS condition, although the answers were shorter, the answers were correct in four out of the five groups, but were less reflective. Three groups used in their answer the correct conceptualization of systematic sampling and stratification. It was also interesting to see that three groups made the sampling method visible in the drawing tool in the same way the sampling method was demonstrated by the teacher during the instruction phase. One group described and drew the sampling method in the correct manner, but did not specify the conceptualization of the method. We found that the one group in the TForPS task that only spent 25 minutes on the task, did not provide a correct and satisfying answer and also exhibited off-task behavior (i.e. drawing a funny smiley in the drawing tool which was meant to visualize their sampling method, see Figure 5). Within the other groups within both conditions no off-task behavior was indicated. The answers of the student groups within the TThroughPS condition better described the procedure of the sampling method they proposed, yet only one student group described their method correctly as stratified sampling. No group used the drawing tool to make visible their method, yet, as depicted in Figure 2, two groups had actively used the provided white boards to support their reasoning. One dome camera captured the student-student interactions of one group in both conditions, the results of these analyses will be presented at the conference.
Based on previous literature indicating that instruction after problem solving provides more opportunities for interactions between students and teacher, both the number and duration of the interactions in both instruction phases were analyzed and compared. Again, we found remarkable differences between both instruction phases. In the instruction phase within the TforPS only 3 teacher-students interactions were observed which only lasted 2 minutes, two were initiated by a question of the lecturer, one was initiated by a clarifying question of a student. The interactions included 2 different students. During the course, the lecturer asked 4 more questions (e.g. is everything clear?), but students did not respond. This was not the case in the instruction phase in the TthroughPS condition. In this condition, 6 student-teacher interactions were identified and the interaction time lasted 10 minutes. The interactions included 6 different students and the discussions dealt about the different strategies they applied during their group work.

Regarding the psycho-physiological data of the students, no significant differences in skin conductance level between the two conditions on the one hand, neither between the problem-solving and instruction phase on the other hand, were found. It is however still possible that different students experience a certain part of the course differently than others. Therefore, the covariance between the measured stress and the level of stress indicated in the self-report will further be explored. The teacher’s psycho-physiological data depicted in Figure 6 shows that the teacher experienced more stress or arousal during instruction phase within the TthroughPS condition. This seems logical as it was the first time the teacher gave her course in this way and the intervention expected her to build upon the student-generate solutions by comparing and contrasting them with the correct solution, which takes the instructor out of the comfort zone. During the problem-solving phase, the teacher did not interfere, which is reflected in the data as well. After the intervention, the teacher remarked that it was not easy to refer to students-answers in the instruction phase and she reflected about the need of a co-teaching context.

Conclusion and discussion
This study is based on considerable research on teaching problem solving in classrooms and helps to better understand and solve the remaining questions about this complex form of activity. The field stresses that there is much more than the two extremes in teaching and learning, i.e. pure direction instruction and unguided discovery learning, however more research is needed to provide evidence-based design guidelines regarding these problem solving design within the middle position (cfr. productive failure and productive success) to achieve optimal learning for all students.

More specific, this quasi-experimental study focused on the order of the collaborative problem solving task. While the first research question focused on the effect on the outcome variables, the second research questions focused on the process variables including the teaching mechanisms during the instruction phase and the learning mechanisms during the problem-solving phase. Although we did not find significant effects regarding the conceptual understanding both on the short and longer term, the results of the affective outcome variables and the qualitative analysis are consistent with the growing body of evidence that generating solutions
to novel concepts and problems prior to instruction can improve learning and teaching compared to students who received instruction first (DeCaro & Rittle-Johnson, 2012; Kapur 2013). The analysis of the quality of answers revealed that starting with problem solving prompts activation of prior knowledge and idea generation from different perspectives.

Although this is a rather small intervention study on a short term, this study complements to the field of the Learning Sciences by bringing together multiple streams of data from both a qualitative and quantitative perspective to fully understand the learning and teaching mechanisms during different approaches towards problem solving and instruction. Taking this zoomed-in perspective gave us the opportunity to capture the interaction processes between the several actors (students, groups and the teacher) and sources (technology, peers and the teacher). The additional analysis of the group work will moreover give us better insight in the collaborative process and the way different students seek for help and or use the provided support which is offered in the task environment. We only started to use psychophysiological data within this educational context and the results within this study did not result in significant differences in the student data, however, we found a considerable trend in the teacher’s data; we will further explore these data as we believe that these data can tell us more detailed information about the affective states of students during teaching and learning in problem-solving contexts. A replication study is currently deployed in secondary education (grade 11 and 12) including the same measurements on a larger scale, including more students and spread over a longer time (4 times 50 minutes instead of 2 times 50 minutes). The first part of the study will be comparable with this design including structured problem solving by means of the web-based inquiry science environment. In the second half of the intervention, however, the problem solving task will be more open and less guided compared to the first one. Consequently, this experimental condition will compare problem solving with a higher chance to lead to success followed or advanced by instruction and problem solving with a higher chance to lead to failure followed or advanced by instruction. According to Kapur (2016) such experimental comparison providing a breakdown of results comparing students who achieved problem-solving success with those who did not is lacking in the field. This extended replication study will provide fruitful insight into the complexity of problem solving, which is situated between the two extremes, i.e. pure direction instruction and unguided discovery learning.

References


Experts' Goals and Constraints When Discussing Vaccines With Laypeople on a Facebook Group

Aviv J. Sharon, Technion – Israel Institute of Technology, a.sharon@technion.ac.il
Ayelet Baram-Tsabari, Technion – Israel Institute of Technology, ayelet@technion.ac.il

Abstract: Science education and science communication scholars have dedicated considerable attention to the fact that laypeople must rely on experts to make decisions in everyday life. Direct communication between experts and non-experts has been facilitated by social media. However, little attention has been given to the processes and outcomes of these communication events, and particularly, to the goals and constraints of the experts. Here, we characterize experts' considerations for answering science- and health-related questions on a questions-and-answers (Q&A) Facebook group dedicated to vaccines. Results indicate that experts have a diverse set of goals and constraints, including disseminating knowledge, defending science and calming fears. Constraints include maintaining collegiality and avoiding provision of medical advice. In light of the findings, we propose a concept of experts' "bounded engagement with the public."

Introduction

Recently, the scholarly community has dedicated increasing attention to the ways science literacy (SL) manifests itself in everyday life. SL has been loosely defined as the knowledge useful or valuable for individuals "using science in their lives, interacting with science information, and making decisions related to science" (National Academies of Sciences Engineering and Medicine, 2016, p. 3), and as a property of communities and societies as well. Since individuals have a necessarily "bounded understanding of science" (Bromme & Goldman, 2014, p. 60), they must rely on scientific experts, such as scientists and healthcare professionals, to cope with the demands of everyday life in a technologically advanced society.

Based on several strands of evidence, Feinstein (2011) has characterized SL as the ability to "interact with sources of scientific expertise in ways that help [one] achieve [one's] own goals" (p. 180). He also asserted that this is necessarily done from an "outsider's" perspective. Using similar reasoning, scholars have proposed that science education must focus on teaching students how to be intellectually dependent on "insider" experts and how to integrate scientific knowledge when they face a specific practical problem, such as a personal health dilemma or a community's environmental concern (e.g., Aikenhead, 2006; Norris, 1995; Weeth Feinstein, Allen, & Jenkins, 2013).

The sparse research on the usefulness of science literacy in everyday life has focused on lay perspectives, such as those of elderly people planning their heating budgets (Layton, Jenkins, Macgill, & Davey, 1993), or those of parents of children with special needs, such as Down's syndrome (Layton et al., 1993), autism (Weeth Feinstein, 2014) or hearing impairments (Shauli & Baram-Tsabari, 2016). By contrast, less is known about the expert side of lay-expert interactions, science-related or otherwise, especially in online environments, such as social media platforms and questions-and-answers (Q&A) services (Brossard & Scheufele, 2013; Shah, Oh, & Oh, 2009; Stilgoe, Lock, & Wilsdon, 2014). Moreover, existing studies on lay interactions with experts have focused on collective contexts of shared policymaking, while non-policy-related dialogue has remained "under-theorized and under-researched" (Davies, McCallie, Simonsson, Lehr, & Duensing, 2009, p. 338). This dialogue can promote learning among both experts and laypeople, in the social constructivist and socio-cultural senses of the word (Davies, McCallie, Simonsson, Lehr, & Duensing, 2009).

Hence, in this study, we explore dialogue on a controversial, everyday personal health issue in an online community of scientific experts and non-experts, from the perspective of scientific insiders. This is part of a larger study, which will incorporate experiences of non-expert "outsiders" as well. Specifically, we characterize the considerations guiding participation of experts in science and health in a Q&A Facebook group dedicated to vaccines.

Literature review

Public communication of scientific experts and health professionals with laypeople

Scientific experts often view their public communication as a one-way transmission of knowledge, meant to educate knowledge-deficient publics to various ends, e.g., to assuage fears of new technologies (Burchell, 2015; Simis, Madden, Cacciatore, & Yeo, 2016). For example, in Cook, Pieri and Robbins (2004), scientists...
Motivations and objectives of answerers on online Q&A services

Several aspects of Q&A services, such as Quora and Yahoo! Answers, have been studied, including the types of questions asked by asker demographics (Baram-Tsabari, Sethi, Bry, & Yarden, 2006), asker motivations (Morris, Teevan, & Panovich, 2010) and answer quality (Harper, Raban, Rafaeli, & Konstan, 2008). However, few studies have explored Q&A services from the answerer's perspective (Oh, 2012; Shah et al., 2009). The existing studies on answerers are mostly based on questionnaires and user activity data. On Yahoo! Answers, answerers who produced health-related content reported that their main motivations were altruism, enjoyment and a sense of efficacy (Oh, 2012). Harper et al. (2008) observed that in at least one case, a large, active community of both researchers and volunteer users contributed to the quantity and quality of answers. Quantitative data on a Chinese social Q&A service, such as following accounts and "upvotes", suggest that answerers were motivated by peer recognition and social learning (Jin, Li, Zhong, & Zhai, 2015). Additionally, several correlations were found between reported motivations and reported answering strategies. For example, questionnaire data suggests that answerers motivated by efficacy tend to select difficult questions, and a learning motivation correlates with searching for additional information when creating answers (Oh, 2011).

Vaccine hesitancy

We chose to investigate public communication of scientific experts in the context of the public controversy surrounding vaccines, because of the pervasiveness of the phenomenon: Health authorities recommend that all children be vaccinated against several infectious diseases, with only rare exceptions. Health authorities regard routine childhood vaccines as a safe and effective way to prevent and even eradicate disease, yet "[a]t the needle
point, [vaccination] enters the intense social world in which parents and carers seek to help their children flourish” (Leach & Fairhead, 2007, p. 2). Although vaccination coverage rates are high in high-income countries, many parents in these countries hesitate to vaccinate children under their care. In fact, 17% of adults in the World Health Organization's European region consider vaccines unsafe, compared with 13% worldwide (Larson et al., 2016). "Vaccine hesitancy" is a complex decision-making process, that depends, among other variables, on trust in health authorities and in mainstream medicine (Peretti-Watel, Ward, Schulz, Verger, & Larson, 2015). Although scholars have recognized the potential of social media for vaccine promotion (Stockwell & Fiks, 2013; Wilson, Atkinson, & Deeks, 2014), few studies have documented vaccine-related discourse online. While there is some research on question topics sent to expert-based Q&A services (Garcia-Basteiro et al., 2012) and on online discussions about vaccines (Fadda, Allam, & Schulz, 2015; Nicholson & Leask, 2012), none of these studies focused on scientific experts' participation in these discussions. A content analysis of the Israeli Facebook group "Parents Talk about the Polio Vaccine" indicated that experts tended to base their claims on ideas relating to the nature of science and to methods of scientific inquiry more often than non-experts. Additionally, claims, even those made by experts, were seldom supported by evidence (Orr & Baram-Tsabari, 2017; Orr, Baram-Tsabari, & Landsman, 2016).

**Research question**
What considerations guide experts' participation in health and science in a vaccine-related Q&A group on social media?

**Research field**
We conducted this study on the Hebrew-language Facebook group Medabrim Al Hissunim ("Talking about Vaccines"; hereafter TaV), founded in October 2013. TaV had over 28,000 members in October 2017. On the group, askers may pose vaccine-related questions and receive answers from community members at no cost. Most askers are mothers of infants. Answerers are often, but not always, experts in science and medicine, including physicians, nurses and scientists, as well as physicians-in-training and scientists-in-training. A list of expert names and professional titles is available through the group description. According to group rules, non-experts in science and medicine are permitted to answer too, but only if their answers reflect "the evidence-based scientific consensus." Some non-expert answerers are very active as well. Hence, the group can be considered a hybrid between a "community-based" Q&A service and an "expert-based" Q&A service (Shah, 2017). TaV has a "public" privacy setting, meaning that its content is visible to non-members. It is operated by a small, volunteer-based Israeli non-profit organization named MiDa'at (derived from the Hebrew term Haskama MiDa'at, "informed consent"). The organization advocates compliance with the vaccination schedule recommended by national public health officials. MiDa'at and TaV were founded in the wake of Israel's 2013 Polio crisis, continuing and expanding the activity which took place in a previously established pro-vaccine Facebook group, described in Rubin et al. (2016), Orr et al. (2016) and Orr et al. (2017).

**Methodology**
Interviews were conducted with ten TaV answerers as part of a larger multiple case study. Two were physicians; one specialized in family medicine and the other was a pediatrician. Another answerer was a physician-in-training. Two answerers were nurses, and five had completed graduate-level research training in biomedical or medical sciences, or were currently in such training (Table 1). Snowball sampling was used, with "snowball criteria" of being an expert in science or health with at least six months' experience answering questions on the group.

**Table 1: Demographic details of interviewees**

<table>
<thead>
<tr>
<th>Pseudonym</th>
<th>Gender</th>
<th>Professional Background</th>
<th>Interview Length (Minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dan</td>
<td>Man</td>
<td>Physician-in-training</td>
<td>105</td>
</tr>
<tr>
<td>Hila</td>
<td>Woman</td>
<td>Ph.D. student in a biological/medical field</td>
<td>37</td>
</tr>
<tr>
<td>Matan</td>
<td>Man</td>
<td>Physician</td>
<td>48</td>
</tr>
<tr>
<td>Vered</td>
<td>Woman</td>
<td>Nurse</td>
<td>39</td>
</tr>
<tr>
<td>Yokhi</td>
<td>Woman</td>
<td>Physician</td>
<td>74</td>
</tr>
<tr>
<td>Ma'ayan</td>
<td>Woman</td>
<td>Nurse</td>
<td>69</td>
</tr>
<tr>
<td>Shlomo</td>
<td>Man</td>
<td>Ph.D. in a biological/medical field</td>
<td>37</td>
</tr>
</tbody>
</table>

ICLS 2018 Proceedings 418 © ISLS
Approximately 30- to 60-minute-long interviews were conducted with each participant, using a two-part protocol. The first part of the interview was a semi-structured interview focusing on the interviewee's motivations for participating in TaV and on their perceptions of askers. Some questions were: "In your opinion, what prompts askers to ask questions in the group?" and "How do you think that askers decide whether to trust you?". The second part was a modified "stimulated recall interview" (Lyle, 2003; Shubert & Meredith, 2015), which can also be considered a modified "reconstruction interview" (Reich & Barnoy, 2016). In this part, study participants were given print-outs of five to six TaV threads they had participated in during the 6 months preceding the interview. They were then asked to read each print-out and explain how they interpreted each question and what considerations they had when composing their answer(s). Interviews were conducted at the participants' homes, workplaces or universities, at each one's preference. The interview with Hila was transcribed by the first author and all others were transcribed by a professional service. All transcripts were coded inductively in MAXQDA by the first author. Approval was obtained from the authors’ institutional review board (approval number 2016 – 19) and from the MiDa’at board of directors.

**Results**

Experts used a diverse set of considerations in their online communication about vaccines. Goals included disseminating knowledge about the vaccines and the diseases they help prevent, defending science against unfounded rumors and conspiracy theories, and calming fears of parents. At the same time, answerers were mindful of two salient constraints: maintaining collegiality with their health professional peers and avoiding provision of medical advice for individuals, even though askers' questions are often phrased in individual terms (Table 2).

### Table 2: Goals and constraints of TaV experts

<table>
<thead>
<tr>
<th>Theme</th>
<th>Prevalence</th>
<th>Example quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disseminating knowledge (Goal)</td>
<td>High</td>
<td>&quot;[The mother] asked, 'does a two- or three-week delay in the vaccines for four-month-olds hurt the efficacy of the vaccine in any way'. I answered that there was no problem at all in terms of efficacy. The only consequences had to do with delaying the protection that vaccines give.&quot; (Hila)</td>
</tr>
<tr>
<td>Defending science (Goal)</td>
<td>Medium</td>
<td>&quot;I wrote that one of the rumors that annoyed me the most […] was about immunological memory. […] People keep saying as if it were a fact: 'Infants have no immunological memory. Obviously.' And that's so wrong. It's so easy to disprove.&quot; (Hila)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In response to a mother who asked whether MMRV, a live attenuated vaccine, provided only community immunity, rather than protecting the vaccinated child. The mother asked the same question in an anti-vax Facebook group: &quot;So I took this mix-up that she made and made a snide remark: Pay attention to the people giving you answers elsewhere. 'Whoever told you that […] doesn't know what they're talking about. Really, I suggest you always ask whoever's giving you answers about their education and training and what they're basing their answer upon.'&quot; (Dan)</td>
</tr>
<tr>
<td>Calming fears (Goal)</td>
<td>Medium</td>
<td>In response to a mother who observed her child talking less often, and suspected that vaccination caused a decline in language development: &quot;I started from the point: 'You did the right thing.' Because she wrote that she may have done damage with the vaccine. 'On the contrary, you did the right thing. You protected your child. Vaccines have nothing to do with language development.'&quot; (Dan)</td>
</tr>
</tbody>
</table>
|                              |            | "A lot of the time, [if askers are in emotional turmoil] I don't answer. I let [other] parents respond from their perspective as parents. And if I do, many times I say: 'Look, I'm doing a PhD in [a biomedical field –  

ICLS 2018 Proceedings

© ISLS
redacted], I took many classes, I read a lot of papers and I vaccinate my children.' Sometimes that helps too. It reassures." (Hila)

| Maintaining collegiality (Constraint) | Medium | "If there's a measles outbreak [in Europe] and people say, 'the nurse wouldn't give a vaccine,' you know, I [maintain] collegial[ity] to the [staff of the] well-baby clinics. [...] I give the rationale and why we act the way we do, and what the rules say and who guides us." (Vered)

"Someone I know who lives in [town – redacted] asked a question [on TaV] about a vaccine for her one-year-old child, who had a very very long fever, and we sent her to a doctor and the doctor told her things. [...] I was tempted to ask who the doctor was who told her this nonsense, because he said it was from the vaccine even though the child had sores in the mouth and all the signs of foot and mouth disease, that can cause a really long fever, but it clearly wasn't an adverse reaction to the vaccine. [...]SO WHAT DID YOU DO?] Nothing. [...] It's none of my business. I need to let her professionals do their work even if I think they're wrong." (Yokhi)

| Avoiding provision of medical advice (Constraint) | Medium | "Somebody once told me – one of our supervisors here [at the well-baby clinic] – 'Be careful when you answer [on TaV] because they can sue you. So that made me think more, because, in fact, I don't give individual answers anymore. [...] I don't give individual advice, I write my answer more generally." (Vered)

Discussion

Experts exhibit a complex set of goals and constraints in their communication with askers on TaV. In some ways, findings evoke previous research on public communication of science by scientific experts, but in other ways, they diverge from the literature. For example, answerers report taking a defensive stance against anti-vax activists, like the "defending science" motivation reported by many scientists communicating online. On the other hand, they exhibit a caring and nurturing form of science communication, offering health information and social support for vaccine hesitant parents who specifically ask for it. TaV can be considered a community for on-demand dissemination of scientific information. This communication is a modified form of the deficit model, characterized by being guided by the publics' genuine information needs. This, arguably, makes it more "legitimate" than other dissemination activities (Trench, 2008, p. 130), which have been characterized as "top-down," "hierarchical," "linear, pedagogical and paternalistic" (Bucchi & Trench, 2014, pp. 3-4).

However, the constraints also deserve further reflection. Just as there is discussion of publics' "bounded understanding of science" (Bromme & Goldman, 2014), we propose to introduce the concept of experts' "bounded engagement with the public". Even the most well-intentioned experts are bounded by their own knowledge (including knowledge of the askers' specific context) and by social constraints, such as the relationships that experts must maintain with their peers.

Further research on co-construction of scientific and health knowledge in online communities can take several directions. Firstly, whether answerers achieve their goals vis-à-vis askers remains to be examined empirically through research with askers. Do dissemination efforts result in knowledge gains? Do answerers' efforts to calm askers truly work? Secondly, further research could characterize the competencies needed for experts for meaningful dialogue with laypeople. Do experts learn these competencies "on the job" or do they draw on skills learned beforehand? To what extent can these competencies be taught, and what are some effective ways to do so? Thirdly, the concept of experts' "bounded engagement with the public" raises questions about the nature of that engagement. What kind of lay questions do experts think they can reasonably answer, and how does that compare with laypeople's expectations? Research into these questions can advance understanding of science literacy on the community level, both within technologically-enhanced environments and outside of them. It can also advance the understanding of co-construction of knowledge in online communities.

References


Doing Science With Fidelity to Persons: Instantiations of Caring Participation in Science Practices

Christina (Stina) Krist, University of Illinois at Urbana-Champaign, ckrist@illinois.edu
Enrique A. Suárez, University of Washington, suareze@uw.edu

Abstract: This paper builds on an emerging line of argument that STEM education should be guided by not only conceptual and epistemological goals, but also axiological ones: attention to the moral and ethical underpinnings that guide learning and participation in scientific and engineering practices. We draw specifically on Noddings’ care theory (1984, 1986), which foregrounds concern for collective well-being and growth. In particular, we explore what a caring orientation can look like in the context of doing science through two paradigmatic cases. These cases highlight how values of receptiveness, responsiveness, and relatedness shape students’ social and epistemic practice. These values challenge canonical versions of science practice and invite reflection on the moral underpinnings of learning science.

Rationale and framing
This paper builds on an emerging line of argument that STEM education should be guided by not only conceptual and epistemological goals, but also axiological ones: the moral and ethical underpinnings that guide learning and participation in scientific and engineering practices (Bang et al., 2016; Philip, Gupta, Elby, & Turpen, 2017). We draw specifically on Noddings’ care theory (1984, 1986), which highlights an orientation to concern for collective well-being in relationships, such as those formed in learning environments between students, as a central ethic. In particular, we are interested in what a caring orientation looks like within the context of doing science, as members of a community attend simultaneously to the conceptual and epistemic dimensions of practice and to the quality of their relationships. As a result of this orientation, our hope is the creation of science learning environments where students are not only motivated to develop powerful scientific explanations, but also to attend to and facilitate opportunities for their peers to fully participate in the process of co-constructing knowledge. This axiological emphasis on foregrounding mutual care is a shift in focus from the often taken-for-granted goals of developing ideas with explanatory power without regards to how people interact with each other to develop those ideas (Roth & Barton, 2004). Attending to relationships allows us to identify ways that students participate and learn that differ from the prevailing attention to the conceptual and epistemic dimensions of learning. Moreover, this approach to understanding learning and participation can provide a more nuanced analytical toolkit for describing productive interactions between students, even if those interactions do not fit disciplinary standards for what is conceptually and epistemically productive. We argue that the goal for fostering a caring orientation is not in competition with the conceptual and epistemological goals of science learning; but that shifting our values to focus on the development of the person does have implications for what we count as “good science practice.”

In this paper, we present an interpretation of Noddings’ work that directs our attention to the goal of developing students who demonstrate care for mutual, collective growth in science understandings and epistemic practices. We then explore the implications of this argument by presenting two illustrative cases that demonstrate what science practice can look like when students orient towards caring for mutual growth. We argue that students in these cases attended to their peers’ growth by creating opportunities for one another to participate in substantive epistemic work. We discuss the implications of considering an ethic of caring in science learning environments for researchers and educators, emphasizing the importance of attending to and constructing the environment’s axiological commitments. Moreover, we offer implications of this work for the Learning Sciences, highlighting the need to raise awareness about the moral and ethical underpinnings of all types of learning environments.

Current goals of U.S. science education
Research in science education has long prioritized understanding how learners make sense of and co-construct knowledge about the natural world, advocating for learning environments that support students’ deep sensemaking around core disciplinary content (e.g., Brown & Campione, 1996). Current reform efforts in US science education aim to support students’ participation in science practices (NGSS Lead States, 2013; NRC, 2012). This emphasis is drawn from conceptualizing science as practice (e.g., Pickering, 2010) and conceptualizing learning as rooted in social interaction and collaborative knowledge construction (Greeno, 2006; Rogoff, 1993). Foundational to both perspectives is the idea that social interactions matter for the fundamental doing and learning of science.

Following from this foundational assumption, research has aimed to understand the nature of student interactions and their impact on learning, including a focus on understanding how students engage in epistemic
practices of science (e.g., Engle & Conant, 2002; Kelly, 2006), on social discourse practices (e.g., Lemke, 2001) and on processes of disciplinary identity formation (e.g., Brickhouse, Lowery, & Schultz, 2000). Though these strands are complementary and mutually supportive, they have each tended to frame social processes as in service of epistemological goals. We aim to question the ethics of this pattern and to consider alternative possibilities for thinking about the relationship between social interactions and disciplinary work.

Interactions guided by fidelity to knowledge vs. fidelity to persons

We begin by considering how and why students might think they are interacting with each other, in terms of the nature of the interpersonal relationship underlying those interactions. For example, are students arguing with each other to prove whose idea is ultimately right, or are students arguing in order to co-construct a genuinely shared understanding? And what do the nature of their interactions communicate about what counts as growth?

Often, work in science education implicitly assumes that what counts as “growth” is ultimately defined and evaluated by convergence on canonical disciplinary understanding. We do not refute the importance of this point, but we want to point out a way that it is insufficient. When growth is framed this way, there is a lack of attention to the means for getting to that disciplinary understanding. And, despite attention to the development of epistemic and social practices in science classrooms, typical US science classrooms promote an individualistic and competitive ethos for what it means to learn and be competent in schooling (Archer et al., 2017; Carlone, Scott, & Lowder, 2014). These values are rooted in an individualist paradigm characteristic of Western Science, where the objective individual discovers the natural world as is (Harding, 1992). Growth, in this sense, is the “uncovering” of truth. Actions that promote this kind of growth embody a fidelity to knowledge. This orientation is fundamentally utilitarian, in that it undergirds science learning environments where the ends (i.e., explanatory power) justify the means (i.e., the nature of how members of a community interact to refine the explanatory power). It focuses exclusively on whether the person is committed to abstract principles and propels persons to act out of duty to those principles rather than true commitment to one another (Figure 1a). In classrooms, this plays out when we ask students and teachers to focus their efforts on evaluating and refining explanatory models with attention to fidelity to the epistemic criteria of the discipline, but without attending to the quality of relationships or the process of that model refinement. Other people (e.g., classmates, teachers) are simply the context within which to develop better knowledge.

As an alternative, we draw on Noddings’ work advocating for a fidelity to persons: the commitment of one person to the growth and well-being another, grounded in a caring relationship. This orientation guides interaction and joint work. In terms of pedagogy, Noddings argues that the “development of the whole person is necessarily [teachers’] concern” (1986, p. 498) and that “caring involves promoting the growth of those for whom we care” (ibid, p. 499). Thus, teachers’ pedagogical decisions and strategies should support their students in developing as persons who care and are cared for by each other. While not explicit in her writing, Noddings’ focus on caring and attending to the quality of relationships can also be extended to student-to-student interactions: when students are epistemic agents, they are responsible for shaping and supporting each other’s learning. In other words, students’ interactions should also be guided by a fidelity to persons.

Importantly, fidelity to persons is not a replacement for fidelity to knowledge. Instead, orienting to fidelity to persons brings a mutual attention to both the epistemological and social nature of interactions: rather than using people to get better ideas, it uses the context of co-constructing ideas to develop ethical, caring relationships. It is then through interactions based in these relationships that genuinely shared epistemological commitments and consensus ideas can emerge (Figure 1b).

Figure 1. (a) Fidelity to knowledge foregrounds epistemic goals (idea creation). (b) Fidelity to persons leverages both social and epistemic goals to foreground caring relationships.

As an alternative, we draw on Noddings’ work advocating for a fidelity to persons: the commitment of one person to the growth and well-being another, grounded in a caring relationship. This orientation guides interaction and joint work. In terms of pedagogy, Noddings argues that the “development of the whole person is necessarily [teachers’] concern” (1986, p. 498) and that “caring involves promoting the growth of those for whom we care” (ibid, p. 499). Thus, teachers’ pedagogical decisions and strategies should support their students in developing as persons who care and are cared for by each other. While not explicit in her writing, Noddings’ focus on caring and attending to the quality of relationships can also be extended to student-to-student interactions: when students are epistemic agents, they are responsible for shaping and supporting each other’s learning. In other words, students’ interactions should also be guided by a fidelity to persons.

Importantly, fidelity to persons is not a replacement for fidelity to knowledge. Instead, orienting to fidelity to persons brings a mutual attention to both the epistemological and social nature of interactions: rather than using people to get better ideas, it uses the context of co-constructing ideas to develop ethical, caring relationships. It is then through interactions based in these relationships that genuinely shared epistemological commitments and consensus ideas can emerge (Figure 1b).
How might the object of fidelity shape disciplinary work?
We argue that one’s axiological orientation to fidelity—to what the object of fidelity is—has implications for the social and epistemic practices of science. Different objects of fidelity (e.g., getting it right, inviting others into conversation) determine how growth is defined and what interactions are prioritized. For example, participation guided by a fidelity to knowledge can create situations where the priority becomes constructing a coherent explanation, under the assumption that reaching that explanation will be most beneficial for the group regardless of who contributed to that process or how. Alternatively, participation guided by a fidelity to persons can create spaces where most, if not all, students can participate in the sensemaking process, with growth reflected in the extent to which students interact with each other with receptivity and responsiveness.

Building from these theorized possibilities, we seek to explore: what might it look like for an orientation to fidelity of persons to guide the social and epistemic activity of students doing science? We explore this question in two cases, both instances of disagreement that arose in student-to-student conversation around a model of a scientific phenomenon. We chose to look at moments of disagreement because they are when the social and epistemic dimensions of science practice can be in tension: one could choose to abandon, for example, an epistemological commitment to explanatory coherence in order to keep peace with a peer. Alternatively, one could abandon a social commitment to reciprocal relationship in order to preserve the quality of an idea. In other words, it is in these moments of disagreement that the guiding axiological commitments become visible in the nature of students’ interactions. We present two cases that we argue each illustrate a form of fidelity to persons in which students maintain both social and epistemological commitments in a mutually reinforcing way.

Methodological approach: Identifying and analyzing paradigmatic cases
The two cases presented below come from two unrelated research projects. One comes from a corpus of video data collected in an out-of-school science program hosted in a library for elementary-aged bilingual children (ages 7-9) that was focused on leveraging students’ language resources in scientific sensemaking. The second comes from a corpus of video data collected in middle school (ages 11-14) science classrooms aiming to support students’ meaningful participation in science practices. We consider these to be paradigmatic cases (Mills, Durepos, & Wiebe, 2010): we use them to illustrate and explore the phenomenon of students demonstrating fidelity to persons in building science knowledge. In addition, we present cases from two distinct settings to demonstrate replication of the phenomenon across contexts (Yin, 2013).

We identified these cases through independent analyses. Suárez conducted discourse analysis of video data from the out-of-school setting looking for markers of disagreement (Suárez, 2017). The out-of-school case presented here was identified because of its distinctive markers for attenuation of conflict and invitation into epistemic practice. This way of engaging in disagreements signaled that students were holding each other’s ideas accountable in ways that did not push peers away. Independently, Krist conducted an analysis aiming to demonstrate how middle school students’ participation in science practices developed over the course of those three years, with a focus on their use of epistemic considerations (Krist, 2016). The middle school case presented here was identified because students used disciplinarily sophisticated epistemic criteria, but the social norms guiding their interactions were markedly different than those in other class periods.

We began looking at these cases side-by-side during conversations on Noddings’ work. We used fidelity to persons as a “sensitizing concept”: an idea that is not yet clearly defined or operationalized, but that “gives the user a general sense of reference and guidance in approaching empirical instances” and “suggest[s] directions along which to look” (Blumer, 1954, p. 7). We brought these cases together as empirical instances that we had been sensitized to through our previous analytic efforts, and we began looking at them together in order to further refine and elaborate what it looks like for a fidelity to persons to play out in science. In particular, as we compared the cases, we looked for indicators of caring occasions: moments when students were participating in ways that demonstrated social and/or epistemic care for one another. We sought to identify in more operational detail the features of their participation that indicated this kind of caring. We drew on science education literature, critical perspectives of science, and Noddings’ descriptors of “responsiveness, receptivity, and relatedness” (Noddings, 1984) to refine the features. The features we identified include:

- **Asking questions that follow ideas**: students are refining their thinking and questioning in response to what others are contributing.
- **Engaging in “idea play”**: students entertain an idea and follow it through to its logical conclusion before evaluating it as right or wrong, or as productive or not.
- **Using data to invite conversation**: students draw on various forms of data to invite dialogic conversation, rather than attempting to “win” an argument, and in contrast to “pseudoargumentation.”
We described how each of these features played out in moment-to-moment participation in analytic memos, paying attention to how they were connected to the social and epistemic dimensions of science practice.

Findings
In each of the cases presented here, students were trying to answer a specific question about natural phenomena and had arrived at a point where there were multiple possible explanatory models on the table. These multiple possibilities created opportunities for students to attend to each other’s reasoning and logic while also orienting towards a group growth characterized by relatedness, receptivity, and responsiveness. In both cases, students asked questions that followed ideas, demonstrating that they had a sense of accountability to ideas and responsiveness as they were carefully attending to and addressing one another. The out-of-school case highlights students engaging in idea play as a form of receptivity to others’ ideas. The middle school case highlights students using data to invite conversation in ways that foregrounded relatedness in that their use of data was in service of supporting mutual growth and understanding. We present each of these cases, highlighting the interconnectedness of these reflections of a fidelity to persons to the social and epistemic dimensions of science practice.

Circuits case: Engaging in “idea play” as a step toward consensus
The circuit case took place in an out-of-school-time learning environment that created opportunities for students to investigate and problematize electrical phenomena. The case presented here occurred during Session 4 of the program, when four students (Yesenia, Elio, Toben, and Grace) built circuits with batteries, wires, and lamps and explained how they thought electricity flowed to light up a lamp. As they presented explanatory models, differences in their ideas emerged. First, Yesenia and Elio stated electricity traveled in a circular motion from end of the battery to the next; Toben and Grace thought electricity left from each side of the battery and met at the lamp. Second, Yesenia and Elio argued that disconnecting the circuit from the battery would make the electricity “freeze” in the wire because there would be nothing to continue pushing it through; Toben and Grace argued that disconnecting the circuit from the battery would make the electricity in the wire return to the battery, leaving the wires empty of any electricity. They spent the next 8 minutes trying to understand and resolve these differences.

The excerpt we analyze here begins with Yesenia asking Toben and Grace what they thought would happen to the electricity as soon as they disconnected the battery from the circuit:

1 Yesenia: How would it (circuit) work? Would it (electricity) freeze? Would it (electricity) still keep running?
2 Instructor: oh it's circular
3 Yesenia: or would it (circuit) turn off?
4 Grace: It (circuit) would turn off.
5 Instructor: It (circuit) would turn off.
6 Yesenia: And what happens to the uh - to the energy?
7 Grace: Um, um, there would - there would - there would be no – no energy, um –
8 Toben: Why?
9 Yesenia: Why? Why wouldn't there be any energy?
10 Grace: It's because if – if it don't – it don't – if you take off the battery um the energy goes to the battery, to the light bulb, or both. And if the wires were connected ((brings index fingers together))
11 the battery uh the electricity could go through here (wire clamps) and make the light bulb worked.
12 Yesenia: But how does the -
13 Toben: Why?
14 Yesenia: the – how does the energy um cross the light bulb and back into
15 the battery, and then do it again and again? How does it cross?
Plate tectonics case: Using data to invite conversation around a discrepancy

The plate tectonics case took place in an 8th grade science classroom in January 2015, at the end of a 3-month Earth Sciences unit. This unit had students develop models of tectonic plate boundaries and interactions (e.g., convergent, divergent, transform, and subduction) in order to explain how the Earth is changing. The last lesson of the unit involved students selecting two to three focal case sites and conjecturing about the tectonic activity that formed the geologic features in that area. For the first two days of the lesson, students researched their case sites in small groups and formed candidate claims. The following three days students presented their claims to
each other using a “fishbowl” discussion format: two to four students who had researched the same sites sat in the center of the room—in the “fishbowl”—while the rest of the students sat around them in a circle. The students in the fishbowl presented their initial claims to each other and had several minutes to ask questions and work on coming to consensus. Then the “audience” could ask questions or make comments to the people in the fishbowl.

For the first “fishbowl” discussion, two students, Benny and Kaatje, discussed how the Andes Mountains formed. Benny claimed that they were formed by oceanic-continental subduction and used his hands, each representing a plate, to show the plates’ interactions. His gestures made it ambiguous as to which way he thought the plates were moving. Kaatje paused, then asked, “Wouldn’t it be convergent then?” They spent several minutes attempting to clarify which way Benny thought the plates were moving. Throughout this exchange, Kaatje was not comfortable with what Benny was proposing, but it is not entirely clear what she was uncomfortable about. As part of one of her responses, Kaatje pointed out a sliver of shallow water in between a trench off the west coast of South America and the coastline (Figure 2a). She said this sliver was troubling to her because if it were subduction, the trench should be exactly along the plate boundary, not offset from the coast.

After about 5 minutes of discussion, the teacher (Mr. M) paused them, affirmed the work they had been doing, and then opened up the “fishbowl” for audience questions. After Mr. M finished, Kaatje nodded to Sara:

Sara: So, in like this packet how we have - it - on the front page it shows, um, the direction of where the plates are moving. (walks over and crouches down next to Kaatje, showing her the packet) You can see that, like, um, right here--

Mr. M interrupted Sara, asking her to project the packet from the document camera so that everyone could see what she was showing Kaatje (Figure 2b). Sara continued, speaking from the front of the room:

Sara: Um, so like, right here, this is where you’re talking about, this is where the Andes are. See how the plate edges is like, a little bit off the coast, so then that’s why there’s like the light blue (pointing to the skinny strip in between the continent and the Nazca plate) and then the trench (pointing to the Nazca plate boundary). So, it’s still like - its’ - this plate is like, like mostly like continental over here, and then it’s oceanic, so um, I lean towards Benny on that (pause; returns to seat). Because of that.

Using their map of plate boundaries, Sara responded directly to Kaatje’s concern about the sliver of shallow water between the trench and the coast, pointing out the skinny strip of South American Plate in between the western edge of the continent and the beginning of the Nazca Plate. She interpreted this slight misalignment to explain the shallow area that Kaatje was concerned about. She also mimicked Mr. M’s language, saying that she “leans towards Benny,” softly stating her agreement with his claim of oceanic-continental subduction.

In her response, Sara was doing sophisticated intellectual work. She listened and attended closely to Kaatje’s concern about the location of the trench. She then pointed to a specific piece of information on a specific map, and she interpreted it in light of another representation (the elevation map). She connected these interpretations back to how they addressed Kaatje’s specific concern, and to the larger conversation that the class was having by clarifying how it was influencing her thinking: she “leans towards Benny.” This phrasing mirrored Mr. M’s language and overall framing of the task as one where their claims are tentative and shiftable.

Throughout this exchange, Sara was also doing sophisticated social work that communicated caring. By crouching down next to Kaatje, she indicated that she genuinely wanted Kaatje to see and understand the information she was sharing with her. In addition, the specificity of her information showed that she had
understood Kaatje and was choosing to respond in terms of Kaatje’s idea, delving in and reacting to it, rather than presenting an alternative argument. Finally, her attenuation of her own stance as a tentative “leaning” maintained their positions as mutual explorers of this idea, rather than positioning of Sara as “right” and Kaatje as “proven wrong.” Taken together, Sara demonstrated fidelity to persons—in this case, to Kaatje—as she interacted with her to try to co-construct a more satisfying explanation.

This interaction invited continued conversation. Mr. M said that Benny or Kaatje should respond to Sara, who said that she “just wanted to hear [Kaatje’s] response.” Kaatje walked up to the board and said:

Kaatje: Oh, I see where you’re going with it but like, you see like how there are some like, how this part like right here (pointing to the shallow strip), how it’s a thinner part, and then it gets thicker in some areas, so like, doesn’t that kind of change the divergence?

Kaatje responded by pointing out that the sliver of plate between the Nazca plate and the coast of South America varied in thickness: it was narrower in some places and wider in others. She asked if that changed the “divergence.” Her statement pointed out another observation that complicated the clean interpretation that Sara had just made from the plate boundary map, which she also presented with attenuation: “like, doesn’t that kind of change the divergence?” Like Sara, Kaatje interacted in a way that demonstrated receptivity to Sara’s idea, responsiveness to the particulars of that idea, and maintained their relationship as one of mutual co-constructors.

Discussion
With these cases, we showed two examples of what making sense of the natural world while guided by a fidelity to persons can look like. In each of these cases, we saw how interactions that communicated receptivity, responsiveness, and relatedness to one another functioned to communicate care for the nature of the relationship and care for the developing idea. These interactions brought learners and their ideas to the table, giving them space and respect to be carefully explored. At the same time, this exploration was not at the expense of disciplinary accountability. In fact, we would argue that the features highlighted in these cases—asking questions that follow ideas, idea play, and use of evidence to invite conversation—parallel the disciplinary practices and work of science (e.g., presenting and evaluating an argument, playing out abstract models and thought experiments, revising ideas based on new evidence and interpretation). Taking on these caring dispositions does not diminish the quality of the epistemic work these learners engaged in, nor the quality of the knowledge products they co-constructed.

At the same time, the features we highlighted push back on some of the norms and values undergirding disciplinary science. An orientation to fidelity of persons emphasizes caring ways of relating to peers and their ideas, a mode of interaction often relegated in disciplinary practice (Barton, 1998; Harding, 2016). In highlighting and advocating for science practice and learning that is guided by a fidelity to persons, we argue that as a field—as learning scientists, as science education researchers, and as teacher educators—we should go beyond a version of epistemic work that focuses exclusively on the quality of ideas and does not attend to the quality of relationships. We acknowledge that a fidelity to knowledge can support epistemic work undergirded by affective, intellectual, and epistemological entanglement, but at what cost? In our current context, the goal and practices for developing caring people is obfuscated by the emphasis on training future members of STEM disciplinary communities. As a field, we have been concerned with getting students to seek mechanisms and to support claims with evidence, often without bothering to attend to how students engage in epistemic work. If the status quo remains, we run the risk of continuing to alienate students who do not identify with and/or would want to engage in epistemic practices that neglect the nature of the relationship between learners, making them feel dispensable.

We call the instantiation of an ethic of caring that values fidelity to persons in science knowledge-building contexts epistemic caring. By highlighting the ways that sensemaking was organized by a fidelity to persons, our goal is to create awareness among STEM educators about the need to interrogate the axiological commitments that guide their research and learning environments. Ultimately, we hope for a STEM education that is axiologically grounded in epistemic caring. Future research should examine other ways this could play out and explore design innovations that support interactions characterized by epistemic caring. An attention to epistemic caring requires us to think deeply about the moral and ethical messages that we are communicating with each instructional and design decision we make, perhaps even forging new practices and structures for interaction that allow these values to be played out (Bang & Vossoughi, 2016).

References


**Acknowledgements**

This research was funded by the National Science Foundation Grant DRL-1020316 to Northwestern University, the Gordon and Betty Moore Foundation Grant No. GBMF 3475 to Tufts University, and by the NAEEd/Spencer Dissertation Fellowship Program. Any opinions, findings, conclusions, or recommendations expressed here are those of the author and do not reflect the official views of the organizations above. In addition, Krist would like to thank “Mr. M” for his tireless commitment to care for persons and ideas; and Ayush Gupta, Anita Sanyal, and the Students Doing Science group at Tufts University for conversations that made this analysis visible. Suárez would like to thank the Sunrise Public Library for hosting and supporting the out-of-school time science program for immigrant youth, as well as Yesenia, Elio, Toben, Grace, and their families for giving permission to share their stories with a broader audience.
Representations of Progress in a Learning Community
Curriculum for Grade 12 Biology

Alisa Acosta, University of Toronto, alisa.acosta@utoronto.ca
James D. Slotta, Boston College, slotta@bc.edu

Abstract: This paper reports on the design and implementation of several student- and teacher-facing learning analytics representations within a blended learning community curriculum for Grade 12 Biology. Using a custom designed technology environment called CKBiology, these representations captured the real-time progress of the learning community at three levels of granularity: Individual students, small groups, and whole class. Our results focus on the perspectives of students and teachers, triangulating data from a student questionnaire, teacher interview, and CKBiology log files to identify how different representations of progress contributed to awareness, motivation, and the overall practices of the learning community. Grounded in the theoretical model of Knowledge Community and Inquiry, this work seeks to strengthen connections between learning analytics research and the learning sciences.

Introduction

Learning communities are characterized by a culture of learning wherein all participants are involved in a collective effort of understanding (Bielaczyc & Collins, 1999). As Kling and Courtright (2003) observe, “developing a group into a community is a major accomplishment that requires special processes and practices, and the experience is often both frustrating and satisfying for the participants” (p. 221). One prominent challenge in adopting learning community approaches is that of assessment (van Aalst & Chan, 2007). In contrast to traditional forms of instruction, wherein the teacher has sole authority over the assessment of students’ work, learning communities provide students with a greater level of agency, allowing them to “develop ways to assess their own progress and work with others to assess the community’s progress” (Bielaczyc & Collins, 1999, p. 272). Thus, in a learning community curriculum the activity designs must clearly articulate the learning processes, making them visible and accessible for assessment. Furthermore, because learning communities focus on both individual and collective aspects of knowledge production, assessment in these contexts must serve the dual function of both measuring and scaffolding learning, producing a “feedforward effect” that serves to catalyze the development of new knowledge (Scardamalia & Bereiter, 2006; van Aalst & Chan, 2007).

This paper reports on the design and implementation of several student- and teacher-facing learning analytic representations within a learning community curriculum for Grade 12 Biology. Using a custom designed technology environment called CKBiology, these representations captured real-time progress of the community at three levels of granularity: Individual students, small groups, and whole class. Grounded in the theoretical model of Knowledge Community and Inquiry (KCI; Slotta, 2014), this work responds to two of the challenge areas identified by Ferguson (2012) concerning learning analytics: 1) Building strong connections to the learning sciences, and 2) focusing on the perspectives of learners. In this study, we triangulate data from a student questionnaire, teacher interview, and CKBiology log files to respond to the following research questions:

1. What forms of representation allow students and the teacher to perceive progress (or gaps in progress) within a learning community?
2. To what extent do these representations motivate students to contribute to the learning community?
3. How are these representations used by the teacher in orchestrating the learning community?

The goals of this research are closely aligned with those identified by Buckingham Shum and Crick (2016) concerning learning analytics for formative assessment of 21st century competencies: “to forge new links from the body of educational/learning sciences research—which typically clarifies the nature of the phenomena under question using representations and language for researchers—to documenting how data, algorithms, code, and user interfaces come together through coherent design in order to automate such analyses, providing actionable insight for the educators, students, and other stakeholders who constitute the learning system” (p. 8).

Literature review

Knowledge Community and Inquiry (KCI)
For many years, theories on collaborative learning tended to focus on how participating in a group would affect an individual’s performance (Stahl, 2015). However, in the late 1980s two programs of research emerged that situated groups of learners within a broader community level: Fostering Communities of Learners (FCL; Brown & Campione, 1994) and Knowledge Building (KB; Scardamalia & Bereiter, 2006). FCL and KB differ with respect to the objectives of the community, the centrality of student-generated ideas, and the level of emphasis placed on prescribed learning goals. However, both of these research programs advanced the notion that the activities occurring in school classrooms should mirror those of authentic research communities, incorporating aspects of collective epistemology and community-level knowledge advancement (Brown, 1994; Scardamalia & Bereiter, 2006). Building upon this body of research, author Jim Slotta developed a pedagogical model called Knowledge Community and Inquiry (KCI) as a means of integrating the perspectives of KB and FCL and making learning community approaches more accessible to researchers and practitioners. As in FCL and KB, students in a KCI classroom work together as a community, building upon each other’s knowledge and nurturing a collective epistemology. However, in a departure from KB, an important aspect of KCI is the design of curricular scripts (Fischer, Kollar, Stegmann, & Wecker, 2013) which specify the activity sequences, materials, student groupings, and technology elements that serve to guide the inquiry toward particular learning goals. KCI curriculum designs are guided by five major design principles, each accompanied by a set of epistemological commitments, pedagogical affordances, and technology elements (Slotta, 2014).

Student- and teacher-facing learning analytics
Learning analytics (LA) entails the application of data science techniques, such as probability modeling and data visualization, to educational data in order to generate actionable knowledge to support teaching and learning (Duval, 2011; Siemens, 2012). Because of its origins in online courseware environments, which typically embraced knowledge-transmission modes of pedagogy, a large proportion of LA research maintains a focus on assessment at the level of individual learners, emphasizing individual achievement and accountability (Chen & Zhang, 2016; Schwartz & Arena, 2013). A systematic literature review performed by Schwendimann et al. (2016) revealed that the primary audience for most LA dashboards was course instructors (71%) and that the predominant context was university settings. Furthermore, only 5% of papers reviewed included an explicit theoretical basis for its LA designs (Schwendimann et al., 2016). In a subsequent literature review focusing on student-facing LA dashboards, Jivet et al (2017) revealed that only 26 out of 95 dashboards had a) been empirically evaluated, and b) had any theoretical grounding in the learning sciences. Of those that did, 18 of 26 were rooted in cognitivist theory and promoted competitive, rather than collaborative, learning behaviors (Jivet et al., 2017). While some researchers have begun to apply LA to more collaborative learning scenarios (e.g. Bachour, Kaplan, & Dillenbourg, 2010; Blikstein & Worsley, 2016; Ferguson & Buckingham Shum, 2012; Shafer et al., 2009), many of these studies report on tools and approaches that have been customized for the researchers, often entailing specialized equipment, complex visual outputs, or data formatting requirements that impede adoption by students and teachers (Vatrapu, Teplov, Fujita, & Bull, 2011).

This paper responds to a central challenge in learning analytics research of interpreting and responding to analytic information within the flow of curricular activities. Wise and Vytasek (2017) define a learning analytics implementation design as “the purposeful framing of activity surrounding how analytic tools, data, and reports are taken up and used as part of an educational endeavor” (p. 151). LA implementation designs address questions such as who should have access to particular kinds of LA, why these LA are being consulted, and how the LA can be fed back into the educational processes taking place (Wise & Vytasek, 2017). Such questions can be incorporated into a curricular script, which specifies how and when to constrain particular interactions, the sequence in which activities take place, and the roles and responsibilities of individuals within the learning community (Fischer et al., 2013). Whereas scripting refers to the structuring of activities before they are run, orchestration refers to the process of executing a curricular script once the activity has already begun (Dillenbourg, 2015). Several researchers (e.g. Rodriguez-Triana, Martinez-Monés, Asensio-Pérez, & Dimitriadis, 2015) have recognized that LA can play an important role in supporting students’ and teachers orchestration decision-making throughout the enactment of CSCL scripts. We have developed such a script, including both student- and teacher-facing LA representations of progress, to investigate the research questions above.

Methodology
This study is part of a broader design-based research project, wherein we worked closely with a high school biology teacher to co-design a KCI curriculum and corresponding technology environment called CKBiology. In this paper, we report on data collected during the third design iteration of CKBiology, which entailed one curricular unit of a Grade 12 Biology course, on the topic of Homeostasis, that was implemented in a blended learning environment over a 10-week period during the 2016-2017 academic year.
Research context, participants and sampling

This research was conducted at a university laboratory school in a large urban area. Activities took place within two contexts: (1) In a traditional science classroom with a “bring your own device” (BYOD) policy, and (2) in a technology-enhanced Active Learning Classroom, which was constructed by the school with the explicit aim of fostering productive collaborations between students (see Figure 1b). A purposeful sampling approach was used to select the teacher participant. Selection was based upon the teacher’s prior experience in KCI research as well as her availability to design and implement a KCI curriculum during the 2016-2017 academic year. The students who participated were an incidental sample in that they were those who happened to be assigned to the classes of our co-design teacher in two sections of a Grade 12 Biology course (n=28).

CKBiology activity structure

There were two types of activities in CKBiology: Lessons and review challenge activities. The lesson activities complemented traditional classroom lectures, and were performed by students within their regular science classroom using their own devices. There were eight lesson topics throughout the Homeostasis Unit, which were taught over multiple days. Each of these lesson topics was visible on students’ CKBiology home screens (see Figure 1a), with activities enabled sequentially by the teacher as they were taught. Following each lecture, students logged on to CKBiology and selected the corresponding lesson activity, where they were assigned three different types of tasks. The first type of task was to define terms or concepts related to that day’s lesson. The list of terms associated with a given lesson was established in advance by the co-design team based on the learning goals for the lesson. Concepts to be defined were divvied up evenly among students in the class. Students’ definitions for these terms were contributed to the community knowledge base in the form of text-based notes with optional images (see Figure 2). The second type of task was to identify relationships between terms or concepts in the knowledge base. Within the CKBiology interface, students were presented with two terms separated by a drop-down list of relationship types. In this case, there was actually a “correct relationship” between each pair of terms, established in advance by the co-design team and programmed into the software. If a student chose the correct relationship, a line would appear connecting the two terms in the knowledge base. The relationship would also appear as a sentence within each note involved in the relationship. For example, the sentence “lysozyme is a type of antimicrobial protein” would appear in both the “lysozyme” note and the “antimicrobial protein” note. The third and final task was to peer review or “vet” definitions that had been submitted by other students in the learning community. Within the CKBiology interface, students were presented with an anonymized definition followed by the prompt: “Is this explanation complete and correct?” If the student responded “yes” to this prompt, that student’s name would be appended to the note along with the statement “This explanation is complete and correct.” If the student responded “no” to the prompt, a text box and image uploader would appear beneath the original note, and the student would be asked to add any new ideas and/or corrected information. Any additional information entered by the student would be appended to the original note along with the student’s name. Subsequent vets performed on that note would also include this appended information.

Following the lessons, there were two CKBiology review challenge activities, completed by small groups of students within the Active Learning Classroom, whose purpose was to help students apply their knowledge to “real-world” inquiry problems. In the first review challenge activity, students selected an area of specialization (i.e. immunology, endocrinology, nephrology, and neurology) and worked within their specialist groups to solve a series of problems in order to become ‘certified’ in their chosen specialization. In the second review activity, students formed jigsaw groups (i.e. “medical clinics”), consisting of one representative from each specialization. Playing the role of medical practitioners, students had to bring together their diverse expertise in order to diagnose a virtual patient with ambiguous symptoms. This included ordering the appropriate lab tests, explaining the reasoning behind their diagnosis, and identifying possible treatment options—thereby consolidating the knowledge they had acquired throughout the unit. In both review challenge activities, a series of scaffolded questions were presented to students in CKBiology using a shared group display, and responses were entered by different group members using a wireless keyboard.

Materials: Representations of progress in CKBiology

1. Progress Bars. There were three kinds of progress bars used in CKBiology: Individual progress bars, group-level progress bars, and community-level progress bars. Individual progress bars were used for CKBiology lesson activities and were visible to individual students on their home screen beside each lesson activity (see Figure 1a), and at the top of the students’ screens as they progressed through their CKBiology lesson tasks (i.e. explaining terms, identifying relationships, and vetting other students’ definitions). The number of tasks assigned to each
student was calculated by dividing the total number of concepts, relationships, and vets by the total number of students in the class. The knowledge base was considered ‘complete’ when all of the terms had been defined, all of the relationships had been identified, and when each definition had been vetted at least twice. While the number of tasks assigned to a student varied from lesson to lesson, on average students were assigned five explanations, five relationships, and 30 vets per lesson throughout the Homeostasis unit. If a student achieved 100% progress, they would have the option of going “above and beyond” their assigned work to make additional contributions, earning themselves a gold star (described below) and additional progress points above 100%. These additional contributions typically took the form of extra vetting tasks and did not detract from the assigned work of other students. In this sense, no individual student could dominate the knowledge base (e.g., by defining all terms and relationships), and every student was still held accountable for making their fair share of contributions.

Community-level progress bars appeared on students’ home screens immediately to the right of their individual progress bar (Figure 1a). These were expressed as a percentage, with 100% being achieved if all students completed their minimum number of assigned tasks. Students who chose to go “above and beyond” their own assigned work by performing additional vetting tasks could increase community-level progress; however as long as there were students who did not contribute their fair share, community progress could never reach 100%.

Group-level progress bars were used for the review challenge activities only, and were displayed on a large screen at the front of the Active Learning Classroom while students were working (Figure 1b). The group-level progress scores represented the proportion of challenge questions that each group had completed. While all three types of progress bars served to represent the quantity of work that students had completed, other features of CKBiology allowed assessment of the quality of students’ work (e.g. vetting, commenting, and review reports).

Figure 1. (a) Student home screen showing individual progress bars (purple) and community-level progress bars (blue) for each lesson. (b) Group-level progress bars publically displayed in the Active Learning Classroom.

Figure 2. Community knowledge base (left side) and explanation note (right side). Explanations containing incomplete or incorrect information as a result of student vetting are indicated with a yellow dot.

2. Gold stars. The gold star representation was used for CKBiology lesson activities. When students achieved 100% progress for their work on a given lesson, they received the message: “Thank you for completing your submission! Would you like to continue contributing your knowledge to the community?” If the student chose “yes,” a gold star icon would appear beside their individual progress bar (see Figure 1a), and the student would earn additional progress points for each additional task they completed. It was up to the student to decide how much more they wished to contribute—the limiting factor being the availability of notes that had neither been authored nor previously vetted by them. The gold star icon itself was visible only to the individual student,
however the teacher was able to see students with progress scores above 100% from her teacher dashboard (i.e. no gold star was present there).

3. Community knowledge base. For each lesson, students’ contributions to CKBiology were aggregated into a shared community knowledge base, which was visible to all members of the learning community. As shown in Figure 2, concepts and terms with completed definitions appeared in blue and those that had not yet been defined appeared in grey. This feature of the representation enabled all community members to see at a glance where gaps existed in the knowledge base. Clicking on a blue term would open the corresponding note, including the original definition and author, followed by any vetting, images (if present), relationships, and comments (if present). Terms that appeared in grey were un-clickable, and the students assigned to those terms were not directly identified. Within the knowledge base, a yellow dot was used to identify notes that had been deemed ‘incomplete’ or ‘incorrect’ as a result of student vetting. This yellow dot served as a cue to the teacher to take a closer look at these notes and potentially initiate a follow-up discussion to negotiate or improve upon these ideas as a class.

4. Teacher dashboard. For the CKBiology lesson activities, the teacher dashboard displayed each individual student’s progress score, ordered from highest to lowest (see Figure 3a). The teacher could also view the community-level progress bar for each lesson, and could toggle to and from the knowledge base. For the review challenge activities, the teacher dashboard included the group progress overview—the same as was displayed publicly for the students. From the group progress overview screen, the teacher could click on any group’s name and pull up their “review report” (Figure 3b), which displayed the group’s responses in real-time as they progressed through their review challenge questions. These responses were color-coded to correspond to the group member (i.e. specialist) who completed the response. The teacher could then use this information to decide when and where to intervene throughout the activity, and to better tailor her support to each group.

Sources of data
To assess how each of the aforementioned representations was used within the learning community, data was triangulated from the following sources:

Student questionnaire. Students were given a questionnaire using Google Forms, which they completed after their final review challenge activity. The questionnaire consisted of 16 items, which were formatted using a five-point Likert scale ranging from “Strongly Disagree” (1) to “Strongly Agree” (5). Several of the item stems were drawn from the “awareness” and “impact” dimensions of the Evaluation Framework for Learning Analytics (EFLA v4; Scheffel, 2017). For questions referring to each kind of representation (i.e., progress bars, gold stars, etc.), an image of the representation was included immediately preceding the corresponding items. At the end of the questionnaire, students also had the option of submitting open-ended comments. In total, 19 students completed the questionnaire and six students submitted additional comments. Sample questionnaire items for Individual progress bars: This representation makes me aware of my current level of progress; The fact that my level of progress is visible to the teacher motivates me to increase my progress, if necessary. Sample items for the Gold Stars: The ability to earn a gold star motivates me to increase my progress if I’ve already reached 100%. Sample items for Community Progress bars: This representation makes me aware of the level of progress of the whole class; This representation motivates me to contribute further, if below 100%. Sample items for Group Progress bars: This representation allows my group to see when other groups are stuck; This representation motivated me to contribute further, if below 100%. Sample items for the Community Knowledge Base (Concept Map): This representation accurately captures all of the important terms/concepts for a given lesson; This representation makes me aware of any gaps in the knowledge base.

Teacher interview. A semi-structured interview with the teacher was conducted following the final review activity. The interview was structured around four images, which were discussed in turn: (1) The individual student progress overview, (2) the group-level progress overview, (3) review reports, and (4) the
community knowledge base representation. The initial prompt for each of these images was “Within the context of a learning community, how useful was this representation to your practice?” with follow-up questions emerging from the resulting discussion. The interview was audio-recorded and transcribed, and lasted approximately 30 minutes in duration.

**CKBiology log data.** Two types of log data were used for this study: (1) Students’ individual progress scores for each lesson, and (2) students’ gold star earnings. Only data from the Homeostasis Unit was analyzed. Additionally, only the first seven of eight lessons were included in the analysis; the final lesson was excluded due to an adjustment in the code that artificially boosted students’ progress scores.

## Results and discussion

### Progress bars

There were two student questionnaire items included for all progress bar representations: (1) An “awareness” item (i.e. “This representation makes me aware of [my/my group’s/the community’s] level of progress”), and (2) A “motivation” item (i.e. “This representation motivates me to contribute further, if below 100%). The responses to each of these items are shown in Figure 4 below.

![Figure 4. Students’ perceptions of the individual, group, and community-level progress bars with respect to their “Awareness” (left side) and “Motivation” (right side).](image)

We performed a Friedman test to identify significant differences in students’ perceptions of the individual, group, and community-level progress bars with respect to their “awareness” and “motivation” ratings. While the Friedman test did not reveal any significant differences among students’ “awareness” scores, significant differences were identified for students’ “motivation” scores, \( \chi^2(2, N=19) = 8.55, p < .05 \). In order to identify which pairwise comparisons were significant, we performed a post-hoc Conover-Iman test on the “motivation” data, including a Bonferroni correction. Results indicated that there were no significant differences in students’ ratings between the individual and group-level progress bars, however the community-level progress bar was rated significantly lower than both the individual and group-level progress bars (both \( p<0.001 \)). These results suggest that students felt significantly less motivated to make contributions to the learning community when they saw that the community-level progress bar was below 100% than they did when their individual progress bar or their group’s progress bar was below 100%.

Data concerning the context and visibility of these representations provides further insight on student motivation. For example, 79% of students agreed or strongly agreed with the statement, “The fact that my level of progress is visible to the teacher motivates me to increase my progress, if necessary.” This suggests that the teacher’s role as an evaluator of students’ work maintained a heavy influence, even within a context of collective cognitive responsibility (Scardamalia, 2002). The values of a learning community were in direct conflict with students’ inclination to focus on competitive, merit-based aspects of schooling, including university applications.

### Gold stars

Only 42% of students agreed or strongly agreed with the statement, “The ability to earn a gold star motivates me to increase my progress, if I’ve already reached 100%.” The CKBiology log data revealed that an individual student’s gold star-earning behavior did not change very much from lesson to lesson: The students who earned a gold star in Lesson 1 (Group A, \( n=12 \)) tended to remain gold-star earners, while students who did not earn a gold...
star in Lesson 1 (Group B, n=16) tended to remain non-gold star-earners. A Welch’s two sample t-test was performed to compare the difference in the number of gold stars earned by these two groups. Results indicated a significant difference in the mean number of gold stars earned by Group A (M=4.2) and Group B (M=0.12); t = −6.9509, p < .0001. Thus, if a student did not earn a gold star in Lesson 1, they were unlikely to earn a gold star in any of the subsequent lessons. Using the same two groups, we compared students’ mean progress scores for all seven lessons. A Welch’s two sample t-test revealed a significant difference in the mean progress score for Group A (M=119.9) and Group B (M=87.3); t = −3.2741, df = 19.268, p < .01, with students in Group A having a 32.6% higher mean progress score than students in Group B.

Knowledge base representation (concept map)
Sixty-three percent of students agreed or strongly agreed with the statement that the knowledge base accurately captured all of the important terms/concepts for a given lesson. In the “additional comments” field on the questionnaire, one student wrote: “There were many terms included that were only circumstantially related to the unit.” In her interview, the teacher also commented: “I think we need to trim the number of terms because it’s just too many. So, I think we should focus more on the basic ones… But that is not something that we would have known going in. Like, this is something that I am actually reflecting now that I went through it.” The teacher also commented that it would be helpful to have two different kinds of vetting dots—one for when an explanation is incomplete and another for when an explanation is incorrect: “Because many times I went into the yellow dots and there was no conflict. There was just, like…somebody put half the definition and then the second person put the second half of the definition, and then a third person came in and said ‘oh wait a minute, and these are examples of blablabla,’ which I thought was great... And then you can take it up in different ways.”

Teacher dashboard
In her interview, the teacher commented that the lesson progress overview screen was “very useful because it made very clear what was happening.” In using the progress overview screen as part of her workflow, the teacher would look for progress scores that she felt were of concern, and would then delve deeper into those students’ work. For example, if she saw a student with an exceedingly high progress score (e.g. in comparison to other students, or to past behavior), she would check the knowledge base to make sure that the student’s explanations weren’t overly superficial or had been flagged as “incomplete” or “incorrect” by other students. Conversely, if she noticed that a student who was typically a high achiever had a low progress score, she would follow up with the student to see what was happening. Regarding the review challenge reports, the teacher commented: “That was really nice. I like the color-coded because it was easy to follow who was doing what. So, I liked them. It was clear. I’m very visual, I think...colors help me.” The teacher indicated that she would mostly use the review reports to check the answers of groups who claimed they were finished: “So ok, these people are done, so I’m going to go see their answers. Then I would go and check ‘ok, so no - this is not great ‘mmm, this needs to be looked after.’ Then I would go back to them and say, ‘did you consider blahblahblah.’...And that’s how I used it.”

Implications and next steps
This study represented our first effort to infuse KCI curriculum and technology environments with learning analytics. To begin, we chose relatively straightforward functions of progress representation because of their familiarity to students and potential impact on helping the community make decisions in response to this information. The experiences in designing and evaluating these features will guide our future efforts, as we add more ‘hidden’ layers of learning analytics, such as tracking groups’ interests and sending materials or prompts based on contextual information. We can also use more nuanced group analytics to determine when a group might need input from the teacher, and send the teacher an active notification in real time. This active form of tracking and notification contrasts with the ambient role of progress representations employed in the present study.

References


Acknowledgements
This work was supported by the Social Sciences and Humanities Research Council of Canada.
Pathways to Literary Reasoning: Bridging Text and World

Allison H. Hall, University of Illinois at Chicago, ahall33@uic.edu
Susan R. Goldman, University of Illinois at Chicago, sgoldman@uic.edu

Abstract: This is an embedded case study of an 11th grade English literature class with an instructional approach based on goals of apprenticing students into disciplinary literacies of literary reading and reasoning. This work examines how three students approached making sense of literary texts in the context of year-long instruction designed to support students in building knowledge, skills, and practices around literary reasoning. Findings illustrate the complexities of teaching and learning literary reasoning by describing multiple pathways in how students’ own knowledge and experience was used to bridge the text and the world. This work has implications for designing spaces that value a variety of approaches towards literary reading while supporting interpretation of increasingly complex texts.

Introduction

Reading literature can be the impetus for reasoning and understanding that make us more thoughtful about the world and our experience of it (Langer, 2011). Scholes (1985) sees texts as imbued with power and sees learning how to critically interpret texts in the world as a form of empowerment. This approach to literary inquiry involves readers as active participants in interpreting and constructing meaning around what those texts might suggest about the world we inhabit and the way people exist in that world. Reading bridges the world created by the words of the author and the world of the reader, what Rosenblatt (1994) referred to as meaning in the transaction of the text with the reader. A reader’s response to a text is to create a new text (Scholes, 1985), “poem” (Rosenblatt, 1994), or “experience” (Dewey, 1934). This new text is never fixed, as our understandings and responses are continuously being revised as we read, think about, or discuss a text. No two experiences of reading are the same, as each individual brings their own knowledge, experiences, and emotions to their reading. The idea that there is no one single correct interpretation underlies both the complexity of reading literature and the complexity of teaching and learning how to read literature. This stance toward reading may be particularly important during adolescence, a developmental period during which youth wrestle with questions of identity, self, self in relation to others and the world, and their experiences in the world. The overarching issue addressed in this paper concerns variations in approaches adolescent readers adopt as they are introduced to and begin to take up a critical and interpretive orientation toward reading literature. The context of the work was a classroom in which the instructional design intentionally sought to engage students in the practices of literary inquiry. The present paper discusses three students as cases embedded in the same classroom context.

Instructional design for literary inquiry

The design of the learning environment in this case was based on research around what it means to reason about literature and how that might be supported through instructional design. In arguing the specific nature of epistemic cognition in literature, Lee, Goldman, Levine, and Magliano (2016) point out that reasoning about literary texts involves both the aesthetic elements of a text as well as messages about the human condition conveyed by the text, i.e., what the text says as well as how those ideas are conveyed. This corresponds with ideas put forward by Hillocks and Ludlow (1984) in their hierarchical taxonomy of skills for reading literary texts, which has two levels: literal and inferential. The present work is concerned with the inferential level, beginning with simple implied relationships. These are inferred based on the language and readers’ personal knowledge and experiences. Complex implied relationships are more difficult because they are dependent on the coordination of many different details from both the text and background knowledge. These first two levels of inference remain inside the world of the text unlike the two highest levels, author’s generalizations and structural generalizations. These both require the reader to step outside the text. Author’s generalizations ask the reader to think of the literary piece as a whole and make a claim about what the author is trying to say about human nature and the world outside the text. Structural generalizations require the reader to explain how the pieces and parts of the text work together to create certain effects. The reader has to both notice the structure and explain how that structure functions to create meaning and message. These high-level skills require readers to approach texts looking for deeper meanings by attending to the language and structure (Graves & Fredrickson, 1991; Rabinowitz, 1987). In the work discussed here, the central goal of instruction was to support students in developing the range of inferential competencies, but especially author and structural generalizations.
In the 11th grade classroom that is the focus of this study, the teacher designed the curriculum around the themes of gender and power, both highly relevant and important for these 16/17-year-old students. She used two focal novels, *A Thousand Splendid Suns* (ATSS) by Khaled Hosseini and *The Handmaid’s Tale* (HT) by Margaret Atwood, both of which afford analyses of power and control in relations among characters and between people and greater institutional and social forces. Power dynamics in both novels are explicitly governed by gender roles in the respective social and cultural contexts: the first, Afghanistan in the past 40 years and the second, a future dystopian society.

The teacher developed the curriculum for this class by analyzing sources of difficulty in each focal novel and, consistent with backward design (Wiggins & McTighe, 1998), developed an intentional sequencing and coordination of learning objectives, scaffolds, and texts that would support students in building the knowledge, interpretive skills, and confidence needed for literal and inferential reading of the novels (Figure 1). The sequencing provided students with multiple opportunities to learn explicit strategies related to literary interpretation and writing with shorter more familiar types of texts and with a gradual decrease in scaffolding over time (Lee, 2007; Sosa, Hall, Goldman, & Lee, 2016).

The combination of objectives, scaffolds, and texts is depicted in Figure 2. Introductory activities at the beginning of the year and before each focal novel were designed to prepare students to make sense of the focal novels by providing background knowledge and criteria to develop interpretive claims and support them with evidence from the text. Smith and Hillocks (1988) point out that when students discuss concepts present in a literary work before they read it, they develop a deeper understanding of the text. Starting with shorter more accessible texts can allow students to practice examining the concept in familiar contexts and generate criteria for understanding and discussing abstract concepts, such as gender and power. In addition, students need relevant background knowledge about the cultural and historical context of the text in order to make sense of characters and events, especially when that context may be distant from their own. These considerations were fundamental to the design of instruction.

The first four weeks of the year began with reading and discussing several short articles about gender relations in Afghanistan to build relevant background knowledge around the theme and context of the first focal
novel. In addition, they analyzed several images and short literary texts to build criteria for gender and power and to practice applying those criteria for building arguments around those claims. The rest of the semester was spent reading the first novel, exploring gender and power in the relationships among the characters and their sociopolitical, cultural context. At the beginning of the second semester, activities and texts were designed to build background knowledge around the specific genre of dystopian fiction, the genre of the second focal novel. There were several reasons to expect HT to pose greater challenges for the students. First, unlike ATSS, HT is not set in a real place with actual events and situations as its backdrop. Like most dystopian stories, it is based in a future, imaginary world in which the government or other institutional entity has control over the people. In addition, the HT narrative jumps back and forth in time without warning and the “truth” of the story being told by the narrator is never made clear. Dystopian stories often employ a dystopian protagonist who questions the existing system and engages in actions that attempt to change or destroy the system. Finally, dystopian literature also serves as a vehicle for authors to criticize aspects of the current society in which they live. With these characteristics and sources of interpretive complexity in mind, the teacher designed a series of texts and tasks to familiarize students with the genre of dystopian fiction in preparation for reading HT. During the reading of HT, tasks involved questions around gender, various types of power and control, dystopian fiction as a genre, the dangers of indifference, and why the author used so many “grey area” characters (neither good nor bad). Much of the work of the second semester focused students on articulating how the author used language choice and structural elements to convey the messages that students were identifying in the text, i.e., on author and structural generalizations (Hillocks& Ludlow, 1984).

A critical dynamic of the teaching and learning process in this classroom were class discussions interleaved with reading and writing activities. Research on instructional supports for literary reasoning indicates the importance of using class discussions to build understanding (Applebee, Langer, Nystrand, & Gamoran, 2003; Langer, 2011) and explore multiple perspectives on text (Earthman, 1992). The teacher led some of the whole class discussions but sometimes they took the form of student-led Socratic seminars. During Socratic seminars, the teacher sat at her desk in the back of the class as students discussed their own questions with a large circle of other students. Activities involving discussing and writing were repeated throughout the year in various combinations and with each of the texts. They included various types of discussions with partners, small groups, and whole class as well as various forms of writing, most of which was individual, with occasional small group projects.

The case descriptions reported in this paper rely on data from whole class discussions and individual writing. They illustrate three variations in approach toward literary interpretation, reflecting different emphases and different dimensions of reading and responding to literature. In other words, they reflect the epistemology of literature and literary inquiry that places importance on the reasoning that supports claims about the literary works (Lee at al., 2016b).

Methods
This classroom-based case study was situated in the context of a multi-institution collaboration between researchers and practitioners investigating evidence-based argumentation with multiple texts in middle and high school history, science, and literature classrooms. In collaboration with teachers, the project designed instructional modules for building adolescents’ disciplinary reading and reasoning skills through inquiry practices appropriate to and reflective of the specific subject matter discipline. Module designs were based on a set of core constructs and learning goals for each discipline, developed through the collaborative work of the project design teams in literature, science, and history (Goldman, Britt, et al. 2016). This study is situated in an 11th grade class (16/17-year-old students) whose teacher was a participant in the project’s literature design team and was implementing instructional modules for literary reading and reasoning.

The class was located in a high school that is part of a large urban district. The school served a diverse student population (47.9% black, 43.1% Hispanic, 3.8% white, 1.9% Asian, and 3.3% mixed race; 87% free and reduced lunch). The composition of the focal class reflected the general population of the school. The teacher of the focal class for this study, Ms. Edwards, is White and was in her fourth year teaching at that school and her third year of involvement on the project when the data were collected.

Data sources for this study are lesson artifacts, field notes, and video and audio recordings of classroom observations from the second week of class through the end of the academic year. Audio of meetings and debriefings with the teacher, written lesson plans, student written work, and audio of interviews with individual students were used in addition to data from classroom observations. Data analyses took place in three phases:

1. Coding of fieldnotes for learning opportunities presented by instructional design (e.g. participation structure, text used, focus of instruction)
2. Coding of video of whole class discussions and student written work for literary reasoning practices (i.e., type of claim or support)

3. Close examination of learning pathways in literary reasoning for three individual student cases through the construction of case files.

The findings presented here are of three case studies drawing largely on phase three and interviews. In phase three, discussion contributions and written work compiled chronologically into single documents called case files. Interviews with case study students were transcribed and added to their case files. Then each case file was examined closely to see in what ways student reasoning practices (i.e., types of claims and support) changed over time and in relation to the instructional sequence (Figures 1, 2). The interview data served to corroborate findings emerging from the discussions and student writing. Finally, the cases were compared to determine distinctive characteristics in each students’ approach towards literary reasoning and pathway across the year.

Findings

The three cases discussed here illustrate variations in their approaches towards literary interpretation. It should be noted that in the class as a whole the first two analytic phases revealed that the learning evident in the work students were producing in their reading journals, class discussion contributions, and other written products reflected the correlated focus of the teaching. That is, student practices in general aligned with the requirements of tasks across the year. Thus, general progress of individual students was consistent with the increasing demands of the curriculum. Table 1 provides an overview of the variations among the three cases in their participation and approaches towards literary reading and reasoning across the year as well as how their strategies changed as they moved from the first to the second focal novel.

Table 1: Case study students

<table>
<thead>
<tr>
<th>Case</th>
<th>Participation in discussions</th>
<th>Approach toward literary texts</th>
<th>Change in strategies from ATSS to HT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freddie</td>
<td>high</td>
<td>Analytical</td>
<td>Related text more to real word and real world more to text</td>
</tr>
<tr>
<td>Brianna</td>
<td>low</td>
<td>Empathic</td>
<td>Increased personal and emotional reactions to text</td>
</tr>
<tr>
<td>Vincent</td>
<td>mid</td>
<td>Literary</td>
<td>Increased use of analogies and attention to author intentionality</td>
</tr>
</tbody>
</table>

An analytical approach

Freddie, an African American male, was one of the most actively engaged students during whole class discussions, often volunteering to lead Socratic seminars and sometimes getting into heated arguments over interpretations of the texts. During the interviews, Freddie shared that he had always felt English classes were easy for him, that his ability “came natural.” Freddie’s confidence in his ability to interpret literary texts and defend his arguments was evident in his quickly picking up the literary analysis practices emphasized in the instruction at the beginning of the year, including analyzing textual evidence while considering criteria to build interpretive claims. In addition, his explanations backing up his claims consistently took into consideration the context of the novel and the larger messages he believed the author was trying to convey.

Freddie’s pathway across the year in literary reasoning knowledge, practices and skills reflected the changing contexts, tasks, and learning goals of instruction. In the beginning of the year, while reading ATSS, the focus of instruction was on analyzing the text to understand the complex interplay of gender and power in relationships among characters in the novel, or what Hillocks and Ludlow (1984) refer to as complex implied relationships. The outside texts brought in were ones that focused on helping students better understand the real-world context of the novel (i.e., articles reporting on gender issues in Afghanistan). Freddie used knowledge of that context along with his understanding of events and characters in the novel to make and support claims about gender and power relationships among characters. For example, during the first Socratic seminar, most of the students agreed that a character’s wives had more power than the character; however, Freddie, in opposition to that claim, read an excerpt from the book, explicitly brought up issues of power and gender, and invoked criteria as support: the ability to make decisions ("at the end of the day he is a man and he makes the decisions," "he chose them as his wives not the other way around") and the ability to make and spend money ("he is the one with the money," “he is the cinema owner”). These criteria were ones on the public list of criteria for gender and power that had been generated by students earlier in the year. His use of these criteria also took into consideration the particular cultural context of the novel so relied on Freddie’s understanding of the world the characters inhabited. This knowledge included the idea that in Afghanistan men made decisions, including who to marry, and that only men could hold jobs outside the home and build wealth.
Then, during the second semester, as instruction moved to focusing on understanding the larger messages of the dystopian fiction novel, Freddie expanded his strategies for interpreting the text, specifically to discerning and describing similarities between the dystopian fictional world of the book and the history of African Americans in the US. Most of these claims fall into the category of author’s generalizations in Hillocks and Ludlow’s (1984) taxonomy as they reflect a “conception of the human situation as it exists outside the limits of the work” (p. 12). For example, in the final discussion of the semester, after students had read the Historical Notes at the end of the novel, the teacher asked students why Atwood included the historical notes instead of just ending with the story of Gilead, the dystopian society. Freddie responded:

Freddie: The way I kinda saw this was ok, so right after slavery there was sharecropping and then after that there was segregation, when I compared this to it, this was like the aftermath sort of like they are not fully accepting women, but the women are in society but there is still discrimination, like after slavery most White people didn't stop and still thought of African American as like a disgusting race and stuff like that, so I think it is sort of like they still in the aftermath they still consider women, that's why he laughed, because they still consider them like a joke.

This final contribution to discussions around HT expanded on Freddie’s ongoing comparison of the two worlds making the case that the story of the women in the novel directly paralleled the story of African Americans in the US. He used the world to interpret the texts just as he used the text to interpret real world events.

An empathic approach

Brianna, an African American female, expressed early in the year her love of reading, making her stance as a “reader” clear throughout the year both in her contributions to discussions and in her writing. In her mid-year interview, Brianna reinforced this stance by immediately making it clear that she didn’t like writing but loved reading: “that’s about all I like about English or literature… I like mysteries, I love love stories, I like a lot of stuff like so different, like you be so into this one thing that you think is right to you, and then someone is like, nah, it’s like this, and then you like oh my god!” Throughout the year, her discussion comments and writing revealed her openness and emotional investment in literary texts.

Brianna’s approach towards literary reading defined her pathway across the year as her approach and willingness to engage emotionally and personally intensified across time. While reading ATSS, one type of response she had was in expressing hopes for the lives of the characters in the text. In an essay arguing that a male character had more power than his wife, Brianna ended with: “She seem to have no power whatsoever, but maybe that can change for her. Maybe she can have that baby and [her husband] will respect her as the mother of his child.” Her personal reactions were often to express a sort of empathy with or understanding of the situation of the characters. For example, in a passage towards the end of ATSS, the same man compared his first wife to a Volga (an inexpensive, utilitarian Russian-made car) and his second to a Mercedes. In Brianna’s writing about it, she began with an empathic reaction to the comparison: “With this quote, I can say when someone tells you that you are worth less than someone else, it hurts, and that is what [his first wife] may feel.” Brianna’s responses to the characters and events in the text often pulled on her own personal understanding of human emotions to arrive at an interpretation of the scene. These emotional responses are a type of identification with the characters in the text as she imagined herself in the position of the character in the novel (Mar, Oatley, Djikic, & Mullin, 2011). Identifying with the emotional experiences of characters was often a way for her to enter the world of the literary texts and a basis for constructing interpretations.

Brianna’s personal and emotional approach towards text seemed to be accentuated as she read HT. This is evident in an early writing on HT, in which she compared women’s safety and fear of being attacked by men to the present day, relating the events of the novel to her own experience: “You can say that’s how women are today. People tell women to be safe at night and stay to yourself when outside. I’ve been hearing that since I was a little girl.” Here she drew on her own experience, making a comparison to it and the world of the text, similar to what Miall and Kuiken (2002) call “remembered emotion.” In another writing, she again referred to her personal reaction to and experience of reading HT: “Power seems to be big in dystopian stories, and it’s really cool to see that. Going to a world that is normal, and then going to a nightmare is very hard for people. It’s better to read things that is going out of hand than really going through those things.” Here, her reaction to
the text was empathic, to imagine how it might feel to live inside the world of the text and experience what the characters were experiencing. This imagined entry into the world of the text occurred elsewhere in Brianna’s writing: “If you see this through the eyes of a person in Gilead…” Her ability to imagine different perspectives and empathic reactions to emotional experiences of characters in the text characterized her particular approach towards text. Like Freddie, she connected the world and the text, but her connections were emotional and personal, and the world was her own, rather than one experienced by an entire segment of society. Janssen, Braaksma, and Rijlaarsdam (2006) found “a more personal, subjective engagement with the stories” to be a characteristic of stronger students of literature, so perhaps her intensified engagement over time indicated a growing willingness to engage with texts, particularly with an unfamiliar and difficult text.

A literary approach
Vincent, a Latino male, often participated in whole class discussions, usually providing elaborate explanations of his reasoning. His contributions were often distinctive because of his use of analogies or hypothetical examples to explain his point, as exemplified in his response to the question of what he might do differently when reading a literary text in the future. He said that he would look for hidden meanings in new books:

Vincent: Before I was just like if it says I saw the same red apple on the same windowsill in like Miss O’Leary’s house, there should be a reason why the apple is there, like maybe she is trying to feed a bird, I don’t know, maybe she is trying to scare away demons or maybe it is just figurative language for like how Miss O’Leary’s is, so to speak that she is an apple or she is shaped as an apple, you never know.

This response highlights Vincent’s use of hypotheticals to explain a point, but also highlights his particular awareness of, and attention to, literary devices as a way for authors to communicate layers of meaning. For example, he noted that there might be special significance to the apple and where it was located, that it might be a symbol germane to understanding Miss O’Leary. Vincent’s written arguments and discussion contributions show that he regularly took this type of interpretive stance (Goldman, McCarthy, & Burkett, 2015) towards texts as well as his awareness of authors as intentional crafters of text.

Vincent’s writing around ATSS showed his attention to details in the text and search for layers of meaning based on what he noticed in the language of the text. For example, in his essay on the relationship between two characters in ATSS, he argued that a father had power over his daughter because he was able to control her emotions. In one part of this essay, Vincent pointed out a pattern in the text: “[The father] is always mentioned when [the daughter] shows her affection to [her teacher] on this page. This repetition of how much and why she loves [her father] when Hosseini mentions [her teacher] deeply emphasizes how much [her father] means to her.” Vincent saw this pattern of the author’s never mentioning the daughter’s affection for one character without the other character also being mentioned as the author’s way of indicating the daughter’s strong emotional attachment. Further on, Vincent also noticed rupture in the text regarding the daughter’s happiness depending on her father: “Her body language, as implied by Hosseini, displays her emotions and [her father’s] importance to them. The drastic change in body language if he is late further digs into how [the father’s] power over [the daughter] continues to grow, and the emotions [the father] has most control over is her happiness.” Here, Vincent attributed special significance to the unexpected or distinctive body language of a character. In these examples, Vincent used two rules of notice—the occurrence of patterns of language and ruptures—that authors rely on their readers awareness of. That is, rules of notice can be cues to meaning beyond the literal (Rabinowitz, 1987). Vincent noticed a pattern or strangeness in the text and surmised that the author had purposefully selected this language and in doing so, the author intended to indicate some special significance. As part of instruction, acknowledgement and awareness of author intentionality in language choice would become much more central to class discussions and analysis of text in reading the second novel.

Vincent’s tendency to use elaborate analogies in his reasoning increased significantly during his reading of HT. For example, in one discussion, he used a hypothetical example of riding a roller coaster to counter a classmate’s claim that the narrator had kept her eyes open during a kiss because she did not enjoy it:

Vincent: Let’s say we are on the roller coaster, we are on the roller coaster, do you want to close your eyes on the roller coaster because you actually do not like it, it’s scaring you, or do you want to open you eyes on the roller coaster because you are like this is fun, I love this? […] She kept her eyes open like I don’t want to miss this moment cuz I actually enjoy it.
This analogy to a hypothetical example helped clarify why he disagreed with his classmate’s assessment of the narrator’s motivations in one scene in the novel and was immediately taken up as the student responded with an analogy about closing one’s eyes while eating delicious food. Vincent would often use a more familiar everyday situation to understand a situation or character in the text, which was unlike the analogies Freddie used, which were to real social and historical events and situations.

Vincent also approached text aware of and trying to discern author intentionality. In several of his writings, he focused particularly on Atwood’s role as crafter of the novel. In one assignment, students were asked to choose a passage of their own to analyze with a small group of classmates. In answer to why he chose the passage he did, Vincent wrote:

“We are containers, it’s only the inside of our bodies that are important” […] Atwood’s ability to write with metaphors and craft these insanely creative sentences is immensely impeccable. I am forever respectful of her. “We are containers,” it is a metaphor for the fact that they are empty inside.

Vincent’s answer shows his aesthetic response to Atwood’s writing, his focus on the function of language, and his knowledge of metaphors as a literary device. Metaphors per se were never discussed as part of the instruction in the class, so Vincent was pulling on his own experience with and knowledge of literature in this case to describe what Atwood was doing with language in the passage. In another short writing, Vincent expanded on his vision of Atwood’s intentionality as crafter of this text:

The fact that Offred says this story is a reconstruction is a little marketing scheme I’d say by Atwood. She uses the term as in this is not what was originally there. As in, Gilead was either much worse, or the story was fabricated, but I’d say the much worse aspect. It’s a cheeky little detail that brings more attention to the book. This lets us know that there are parts missing, some that could never be recovered. Of course, Offred is a fictional character but it creates an extra dimension to her story that we weren’t there for. Like I said, cheeky.

His response moved between the world of the text and the idea of it as a fictional creation of the author. In the first part, he wrote as if Gilead actually existed: “Gilead was either much worse, or the story was fabricated” by the narrator. Then, towards the end, he emphasized that the narrator “of course” was only fictional. His explanation of why Atwood had her narrator claim that her story was a reconstruction showed Vincent’s view of authors as people who write books that they want to sell, “a little marketing scheme.” In Vincent’s view, Atwood used the idea of a reconstruction to create a bit of mystery around the world of Gilead to get the reader’s attention, as a sort of “hook.” This excerpt also shows Vincent’s own distinctive use of language in his description of Atwood’s moves as a writer as being “cheeky.” Vincent seemed to have a strong opinion of Atwood as an author.

**Discussion and implications**

The three cases illustrate three qualitatively distinct approaches towards reading and interpreting literary texts, different ways of engaging in the practices of the class, and various pathways through the designed instruction. However, in terms of the learning goals for the year, all three students developed in their critical engagement with literary texts. They all were able to make interpretive claims and support them with textual evidence and appropriate reasoning, which relied on making analogies or connections to the world outside the text. However, the nature of these differed. Freddie most often related the text to his knowledge of history or real-world circumstances. Freddie’s analogies reflected his analytical approach towards text. Brianna, on the other hand, related the text to her own personal, lived experiences, reflecting her emotional, personal response to literature. Vincent would often make analogies to everyday hypothetical situations. This creativity in reasoning echoed his repeated appreciation for creativity in the author’s writing. Interestingly, each student’s trajectory was significantly impacted by the change in texts, tasks, and focus of instruction during the second half of the year. As the texts and tasks increased in complexity and required a greater level of abstraction, all three were able to build on and deepen their particular approaches to meet the increasing challenges.

Each approach reflects an acceptable way of “doing” literary reading. That is, knowledge, experience, and emotion are each legitimate and valuable ways of approaching literary texts. Interestingly, we see each one of the students using primarily one of these resources in their interpretations throughout the year. Each student “read” the same text at the literal level, yet each took a different approach to interpretation. Importantly, each approach was valued and accepted in the classroom context; none were questioned or rejected. Expressing, listening to, and considering multiple perspectives on text is fundamental to the complex nature of literary reading. These cases from the one classroom indicate a classroom community where these epistemic norms and
values are alive and well. The variations in how these three students approached literary interpretation are natural and in some sense predictable, especially for literary reasoning and interpretation where so much depends on who the reader is and how they transact with the text. Indeed, it appears that each of these students found their own way of transacting with the text, seeing the world in the text and the text in the world. Each individual’s world and experience of it is different, as is what each individual reader brings to a text or seeks from a text. These different ways of interacting need to be supported and made manageable by the instruction, while instruction still supports students in being able to read more complex texts and engage in more complex tasks over time. Lee et al. (2016) hold that knowledge in literary reasoning is complex, multidimensional, and contextual. Therefore, designing to support development of that knowledge needs to consider all the complexities and dimensions of readers’ experiences of and approaches towards texts in order to leverage and build on them while also creating a community where students can learn from each other’s perspectives.

References


Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100007 to University of Illinois at Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.
Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads

Emma L. Starr, Joseph M. Reilly, and Bertrand Schneider
ems519@mail.harvard.edu, josephreilly@g.harvard.edu, bertrand_schneider@gse.harvard.edu
Harvard Graduate School of Education

Abstract: This paper describes an empirical study where the productive interactions of small collaborative learning groups in response to two collaboration interventions were evaluated through traditional and multi-modal data collection methods. We asked 42 pairs (N=84) of participants to program a robot to solve a series of mazes. Participants had no prior programming experience, and we used a block-based environment with pre-made functions as well as video tutorials to scaffold the activity. We explored 2 interventions to support their collaboration: a real-time visualization of their verbal contribution and a short verbal explanation of the benefits of collaboration for learning. This paper describes our experimental design, the effect of the interventions, preliminary results from the Kinect sensor, and our future plans to analyze additional sensor data. We conclude by highlighting the importance of capturing and supporting 21st century skills (i.e., collaboration and effective communication) in small groups of students.

Introduction
For decades, the field of CSCL (Computer-Supported Collaborative Learning) has been concerned with promoting socio-constructivist outcomes. The idea that students learn particularly well in social settings is not new (e.g., Vygotsky, 1980; Piaget, 1998). What is new, however, is that we now have unprecedented means to measure and support students’ interaction. On the measurement side, in particular, we can now have access to massive datasets characterizing students’ collaboration and learning processes. Recently, sensing technologies have become increasingly affordable and easy to use, allowing researchers to collect large datasets on students’ interactions in real-world settings, such as classrooms, makerspaces, museums, or other informal learning environments. In the near future, we are envisioning that a combination of qualitative and computational measures will provide us with rich information about the different facets of productive collaborative learning groups. Being able to accurately measure collaborative skills is of primary importance, because our current educational system tends to teach what it can measure. If we can develop innovative and automated ways of capturing those skills, we can pave the way to new forms of formative assessment and new ways of teaching those skills in traditional school curricula.

Theoretical framework
While there is a wealth of theories of collaborative learning, we focus on Roschelle’s (1992) framework of convergent conceptual change. In this framework, collaboration is seen as the process of constructing shared meanings for conversations, concepts, and experiences. It gradually leads to the construction of new meanings and results in conceptual change. From this perspective, markers of collaborative learning are captured through iterative cycles of interactions that converge toward a shared understanding of the task and the concepts taught. For example, researchers in Computer-Supported Collaborative Learning (CSCL) have developed tools to capture the level of transactivity of a dialogue over time (i.e., the extent to which students build on each other’s ideas in small groups; Ward & Litman, 2007). So far, the main tools for capturing these processes are limited to time-consuming qualitative analyses or quantitative methods applied to verbal exchanges (transcripts) and self-reports (surveys, questionnaires). We suggest that new technologies, sensors in particular, can provide researchers with a richer and complementary way of capturing collaborative processes.

Capturing collaboration through Multi-Modal Learning Analytics (MMLA)
Over the past decade, high frequency sensors (such as eye-trackers, motion sensors, wearables) have become affordable and reliable, which opens new doors for capturing students’ multi-modal interactions. They allow educational researchers to collect significantly larger datasets: a sensor typically runs at 30-120Hz and collects various streams of information. The Microsoft Xbox Kinect sensor, for example, can collect information about a person’s body joints (x,y,z coordinates), their facial expressions, and their speech at 30 Hz (i.e., 30 times per second). One can easily define ~100 variables that can be captured from the Kinect sensor. This means 3000 data points per second for one person, which translates to roughly 10 million data points for an hour of data.
collection. Multiply this figure by the number of sensors (eye-trackers, GSR sensors, emotion detection tools, speech features), participants, and studies to get a sense of the possibilities of combining sensors with data mining techniques. The field of Multi-Modal Learning Analytics (MMLA; Blikstein & Worsley, 2016) is about exploiting this new development. Sensors and data mining techniques have both reached a level of maturity that allows researchers to tackle new research questions and develop new educational interventions. MMLA can be used not only to study collaborative learning, but also to support it in innovative ways.

Supporting collaboration through MMLA
The effect of technology-based collaboration interventions on collaboration quality has been examined previously. Bachour, Kaplan, and Dillenbourg (2010) developed an interactive table system known as Reflect, which supported participation for groups of four by presenting a visualization of how much each individual talked during a learning activity. Disparity in collaborator participation can be detrimental to both individual and group experience during a collaborative activity (Salomon & Globerson, 1989). Bachour et al. found that their interactive table intervention helped participants be more aware of their participation levels during the activity, and encouraged groups to be more balanced in participation levels when the participants in the group felt that balanced participation was important to them. This research demonstrates that technological interventions that target participation balance can change the behavior of collaborators during a task. However, their study did not explore the relationship between that behavioral change and other constructs, such as experimenter and participant ratings of collaboration quality, task understanding and completion, nor the learning gains of individual participants. The present study implements a similar visualization as an intervention, while collecting data about these other critical constructs using traditional and multi-modal measurement techniques.

General description of the experiment
In this study, 42 pairs of participants had 30 minutes to program a robot to navigate through a series of increasingly difficult mazes. Two interventions were used to support collaboration: a visualization representing verbal contributions and a short verbal explanation of the benefits of collaboration for learning. Dependent measures included 1) how well participants programmed the robot (task performance), 2) differences from pre-to post-test (learning gains), and 3) the quality of their collaboration (using a validated rating scheme described below). In terms of process variables, we used three kinds of sensors: two mobile eye-trackers, a motion sensor, and two bracelets capturing electrodermal activity. In this paper, we focus on the impact of our two interventions on our dependent measures. More specifically, we predict that each intervention should positively impact task performance, collaboration and learning gains.

Methods

Subjects
Forty-two dyads (N=84) participated in the study, although two groups were dropped from analyses due to experimenter error during data collection. Participants were recruited from the study pool of a laboratory at a university in the northeastern United States. Of the participants in the study, 62.2% were full or part-time students while the rest were members from the surrounding community. Participants ranged in age from 19 to 51 years old with a mean age of 26.7 years, and 60% identified as female. Participants were compensated $20 for the 90-minute session.

Materials
Participant learning was measured with a pre- and post-test learning assessment. The learning assessment consisted of four short answer or fill-in-the-blank questions that assessed their understanding of basic computer science competencies, such as calculating outcomes from a loop and interpreting the purpose of example code (adapted from Brennan & Resnick, 2012; Weintrop & Wilensky, 2015). There were two versions of the learning test with slight differences in question wording and figures.

In addition to the four learning test questions, during the posttest participants filled out a self-assessment of their collaboration experience, adapted from Meier, Spada, and Rummel (2007), and a demographic survey. The self-assessment included six Likert scale questions that asked about features of collaboration, such as task division, time management, and partner respect. The demographic survey asked participants to report their age, gender identity, student status, and educational level.
The task asked participants to use a block-based programming language called Tinker to navigate a simple robot through a series of mazes. The robot was constructed out of a GoGo Board, an open-sourced educational hardware device (Sipitakiat, Blikstein, & Cavallo, 2004). The robot came equipped with proximity sensors on the front, left, and right of the robot and DC motors on each side (Figure 2a). Participants used Tinker and the GoGo Board’s app interface, the GoGo Widget, to learn about and control the robot. Participants were first given a training task to code the robot to go straight across a line approximately two feet in front of them so they could learn about the robot and coding environment. Behavior during this first activity was not included in the analysis. The main activity asked participants to spend thirty minutes attempting to get the robot through three mazes (Figure 1). The first maze consisted of one wall, requiring participants to make one right turn. The second maze consisted of two walls, requiring participants to make seven left and right turns. The final maze required participants to navigate a dead end. Participants were told that the overall goal was to create flexible code so that the robot would be able to solve any maze it encountered.

While participants attempted the task, their collaboration and task behaviors were assessed by the researcher using scales that ranged from negative two to positive two. The dyads’ collaboration was assessed on nine scales, adapted from Meier, Spada, & Rummel (2007): sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation. The task behavior measures included task performance, task understanding, and improvement over time.

Design
This study utilized a two by two between-subjects design. The two independent variables were a speech equity visualization intervention (Figure 2b; inspired from Bachour et al. 2010) and an informational collaboration intervention. The speech equity intervention utilized data from an Xbox Kinect sensor to track how much each participant spoke during the activity. Dyads assigned to receive this visualization saw a tablet that displayed how much each of them had spoken in relation to the other’s speech in the past 30 seconds by presenting rectangles proportional to their speech time. As one participant spoke more than the other, their rectangle grew larger and took up more of the screen. Dyads not assigned to this condition received no visual intervention.

The informational collaboration intervention consisted of a researcher giving participants instructions around collaboration before the study activity began. Dyads assigned to this intervention were reminded that they were expected to collaborate during the learning activity, and were invited to think about how they were collaborating throughout the session. They were told that previous research has found that factors like equity of each partner’s speech time is predictive of the quality of collaboration and learning gains. Dyads not assigned to this condition received no informational intervention.

A quarter of dyads (10 of 40) received no visualization nor the collaboration information (Condition #1), a quarter received only the visualization (#2), a quarter received only the collaboration information (#3), and the final quarter received both the visualization and the collaboration information (#4). These conditions assignments were generated randomly.
Procedure

Each participant first signed an informed consent, and then individually completed the pretest learning assessment. Participants had five minutes to complete the pretest assessment. The two participants worked on different versions of the assessment on separate computers. Participants were then informed that they would be collaborating with one another, and were instructed to introduce themselves to their partner.

A tutorial video was shown that illustrated basic concepts of how to use Tinker to program the robot, such as how to find, insert, and delete code blocks. After the video, participants had five minutes to complete the training task to code the robot to move forward across a line roughly two feet in front of them. If after five minutes they were unable to complete this task, the researcher demonstrated how to accomplish it and explained the rationale behind the code used. If participants did successfully complete the task, the researcher showed them the same code solution that they showed to unsuccessful participants if any differences existed between that solution and the participant’s code. All participants received an explanation of the rationale for this code solution, so that all dyads received the same explanation and code example before progressing.

Participants then saw a second tutorial video, which showed more advanced features of the GoGo Board, including how to use prewritten functions for going forward, left, or right. Participants were also exposed to more complicated code examples, including how to use conditional statements to trigger actions when certain conditions were met, and how to utilize the GoGo Widget. After this, dyads were given 30 minutes to try to code the robot to solve a series of three increasingly complex mazes, as described above and depicted in Figure 1.

While they were completing the series of three mazes, the participants had access to a printed reference sheet that reviewed some of the material presented in the two tutorial videos. Participants also received hints every five minutes from the researcher throughout the thirty-minute session. These hints were the same across all groups, and consisted of reminders of the available tools and suggestions for what code blocks to try to use.

After the 30-minute session, participants had 10 minutes to complete a posttest learning assessment. Participants received the version of the pretest that they did not complete previously as the first section of their posttest. They also responded to questions that asked them to reflect on their experience with the task and about their collaboration experience, and filled out a demographic survey. Finally, participants were debriefed, thanked for their participation, and compensated with $20.

Coding

Dyads’ collaboration behavior and task performance were live-coded by the researcher conducting the session using the scales described above. During the main 30-minute work session, researchers looked for evidence of behaviors that corresponded to one of five levels of each scale, which ranged from good behaviors at positive two and poor behaviors or the absence of good behaviors at negative two. Multiple researchers conducted sessions of the study and thus coded dyads’ behavior. Researchers double coded 20% of the sessions from videos collected during the session, and had an inter-rater reliability of 0.65 (75% agreement).

The four items on the learning test were graded on a zero to three rubric scale to evaluate completeness of answers and understanding of computational thinking skills. These scores were added together to generate total pre- and post-test scores for each participant as well as learning gains. The final code each dyad created was evaluated on a zero to four scale to determine how well the code in abstract could perform the maze solving task. This rubric aligns with the live coding of “Task Understanding” done during the session, acting as a post-hoc assessment to ensure dyads’ final products were fully evaluated.
Multi-modal data
A number of multi-modal sensors were also used to collect data from both participants in each session. Tobii Pro Glasses 2 eye-tracking glasses (https://www.tobiiipro.com/product-listing/tobii-pro-glasses-2) were used to follow where each participant looked throughout the session. The eye-tracking glasses sampled at a rate of 50 Hz, thus generating roughly 90,000 data points per person during the 30-minute session.

An Empatica E4 wrist sensor (https://www.empatica.com/e4-wristband) was used to track several physiological markers from each participant, including electrodermal activity (at 4 Hz), blood volume pulse (at 64 Hz), and XYZ acceleration (at 32 Hz). During the 30-minute session, roughly between 7,200 to 115,200 data points were generated for each participant per measure, depending on the physiological measure’s sampling rate.

Finally, a Kinect was used to track the motions of the dyads. The sampling rate for the Kinect was 30 Hz, generating roughly 54,000 data points during the main session for around 100 different variables. Data from these sensors will be analyzed and presented more thoroughly in future publications.

Results
Assessment of collaboration
Analysis of the coding of dyad collaboration revealed significant differences between the two conditions that received the informational intervention to support collaboration (3&4) versus those that did not (1&2). Dyads in condition 3 scored 7.1 points higher than those in condition 1 (p < 0.001), both of which did not receive the Kinect-based visualization intervention. Dyads in condition 4 scored 4.8 points higher than those in condition 2 (p = 0.03), both of which did receive the Kinect intervention. Differences in collaboration between the conditions that received the Kinect-based visualization intervention (2&4) and those that did not (1&3) were not significant when controlling for the verbal intervention on collaboration.

Participants’ self-reported collaboration scores at the individual level differed significantly from the researchers’ assessment of their collaboration at the dyad level (F = 15.21, p < 0.001) but they are significantly positively correlated (r = 0.43, p = 0.001). Self-reported scores were on average higher for measures of task division, time management, and reciprocal interaction while being lower for reaching consensus, dialog management, and sustaining mutual understanding. Further qualitative analysis and analysis of multi-modal data sources may reveal additional details regarding participants’ self-perceptions of effective collaboration and why they differ from the researchers’ coding. Neither scale differed significantly by individual gender, the gender makeup of the group, or level of education of participants.

Researcher coding of collaboration was significantly positively correlated with the quality of produced tinker code (r = 0.52, p < 0.001) as well as all three performance metrics: task performance (r = 0.35, p <0.001), task understanding (r = 0.53, p < 0.001), and improvement over time (r = 0.54, p < 0.001) (See Figure 3a.)

Learning test
Pairwise comparisons of treatment group means with a Bonferroni correction for multiple comparisons revealed no significant differences between groups’ pretest, posttest, or gains scores by condition on the learning test. All pairwise comparisons were confirmed with a Tukey's honest significance post-hoc test. The two interventions did not target content knowledge gains directly but aimed to increase collaboration within the dyads. While not differing significantly by condition, participants on average gained 19.8 percentage points between the pre- and post-survey (t = 6.18, p < 0.001). This indicates the efficacy of even a short interactive programming lesson targeting computational thinking fundamentals. Learning gains do not differ significantly by individual gender, the gender makeup of the group, or level of education of participants.

Posttest scores on a question that asked participants to explain how to solve a maze were significantly correlated with the quality of the code the dyads wrote (r = 0.26, p = 0.04). A question on interpreting code that used nested conditional statements was significantly correlated with our coding of dyad collaboration (r = 0.33, p < 0.01) as well as the number of mazes dyads completed (r = 0.35, p < 0.005).

Performance metrics and quality of produced code
The mean number of mazes completed by each group (task performance) and improvement over time did not differ significantly by condition. Mean scores on task understanding only significantly differed between Conditions 2 and 3 (p < 0.05) but both interventions differed between those two groups. The quality of the final block-based code dyads produced is significantly correlated with our assessment of the quality of their
collaboration \((r = 0.52, p < 0.001)\), the number of mazes completed \((r = 0.45, p < 0.001)\), their task understanding \((r = 0.45, p < 0.001)\), and their improvement over time \((r = 0.54, p < 0.001)\).

To estimate the relationship between collaboration and code quality, a three-level (participants in dyads in conditions) linear mixed-effects model was fit by residual maximum likelihood methods. On average in the population, a one-unit increase on our rating of collaboration was associated with a 0.071-point increase in the grand mean of the quality of their code \((p < 0.001)\) when controlling for gender and education (Figure 3b.) At a collaboration rating of 0, the expected mean code quality is 1.88 \((p < 0.001)\). To aid interpretation, a 14-point increase in our rating of collaboration corresponded to roughly a one-point increase on our 0-4 code quality scale. Additional model building and discussion of the random effects of the model will be explored in future work.

![Figure 3](image)

**Figure 3.** (a) Correlogram of different performance metrics and ratings of collaboration. All correlations are positive, only significant ones plotted. (b). Linear fit of fixed effects of code quality versus collaboration rating.

**Preliminary results: Movement and talking of dyads**

The total amount of movement across all upper body joints and body parts was calculated for each session as well as the total amount of talking per dyad. The amount of talking was significantly positively correlated with our assessment of the quality of collaboration \((r = 0.45, p = 0.007)\) as was movement of the participants’ left elbows \((r = 0.36, p = 0.03)\) and left hands \((r = 0.36, p = 0.04)\). From observer notes, most participants were right handed and used that hand to manipulate the keyboard and mouse to program the robot. This would leave the left hand more free to gesture. While the intervention related to visualizing participant verbalization did not appear to have a significant effect on quality of collaboration, the strength of the relationship between amount of talking and quality of collaboration suggests this is fruitful area for additional exploration. Future exploration of this data will examine differences in movement and talking within groups and how this affects collaboration as well as determining prototypical postures or gestures that may indicate quality of collaboration.

**Plans to analyze physiological markers, eye tracking, and motion data**

Our main goal is to conduct an in-depth analysis of the sensor data collected during this study. More specifically, our first step will be to design measures of convergence using physiological, gestural, and visual data. For example, it is well-known that Joint Visual Attention (JVA) is a prerequisite for effective collaborations (Schneider et al., 2016). There is also some initial evidence that physiological synchrony is indicative of productive groups (Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2016). Finally, body postures and gestures can be used to identify leadership behaviors - though bodily synchronization has not been found to correlate with collaboration quality (Schneider & Blikstein, 2015). We plan to replicate those findings on our datasets, and combine them to develop multi-modal measures of synchronicity in dyads.

**Discussion**
The purpose of this paper was to explore the effect of two collaboration interventions and the relationship between collaboration quality, task performance, learning gains. The main hypothesis was that the two collaboration interventions, one visual and the other informational, would improve dyads’ collaboration quality. Administration of the interventions was crossed such that an equal number of groups (10 dyads) received neither, one, or both of the interventions. Analysis of researcher’s coding of participants’ collaboration quality found that while collaboration quality improved for both groups that received the informational intervention, there was no significant improvement of collaboration quality for the groups that received the visualization intervention and no significant advantage conferred to the group that experienced both interventions.

The failure of the visualization intervention to influence the collaboration behavior is somewhat surprising. Bachour et al. (2010) implemented a similar visualization intervention that targeted participants’ speech balance as a way to alter collaboration behavior, and found that groups who experienced the visualization intervention had more balanced participation levels when participants reported that they believed balanced participations levels were important. Based on those findings, it might have been predicted that dyads’ who experienced both the informational and visualization interventions would have benefitted the most, as they had the external feedback tool and the importance of participation balance made salient. However, there was no evidence to support this. Possible explanations for this could included the design of the visualization tool. The visualization tool was presented on a relatively small tablet screen on the other side of a table from the participants. It could be that the visualization was not salient enough for participants to be motivated to attend to during the challenging task. Further analysis of the data could examine the eye-tracking of participants to assess how frequently they looked at the visualization and whether that correlates to collaboration quality. Additionally, the tool used in the present study operated on a different timescale than Bachour et al.’s. Rather than showing total participation throughout the session, the visualization was based on the past thirty seconds of speech. Perhaps this timescale was not optimal for altering participant behaviors. Lastly, Bachour et al.’s study did not include any measure of collaboration quality other than speech time balance among participants. It could be that participation balance simply is not an effective predictor of collaboration quality more broadly, and the effect of the informational intervention stems more from the mere reminder of the importance of collaboration than from encouragement of speech equity in particular.

Additional metrics were used to evaluate the efficacy of the two interventions, including task performance and learning gains on the pre- to posttest assessment. However, analyses showed that neither of these metrics were significantly affected by collaboration condition assignment. This is likely due to reasons similar to those listed above. Other reasons could include the challenging nature of the activity that resulted in a performance ceiling effect and that the interventions did not directly addressed the learning goals of the activity.

The present study also aimed to explore the relationship between collaboration quality, learning gains, and task performance. There was evidence that stronger task performance, as measured by final code quality, was significantly associated with researcher’s coding of collaboration quality. Learning gains, as measured by the changes in performance from the pre- to posttest, revealed only modest evidence that certain questions on the test correlated with collaboration quality and task performance. This finding is not surprising, as the research on the relationship between collaboration quality and learning outcomes is mixed (Dillenbourg, Baker, Blaye, & O’Malley, 1996). There was a significant improvement in participant scores from pre- to posttest overall, suggesting that even this relatively short learning activity led to increase in computational thinking skills.

**Conclusion**

While this study was not able to show a clear effect of providing a real-time visualization to support collaboration, it made many other valuable contributions. First, it showed that simple verbal interventions can help participants pay attention to particular aspects of their collaborative behavior (i.e., how much they are talking and how much space they are providing to their partner). Second, it suggested that awareness tools such as the one developed for this study have to be designed differently to impact social interactions (e.g., by being more salient or be used in a setting where users have the mental bandwidth to reflect on their collaborative style). Third, we collected a rich multi-modal dataset that can be used to build proxies for measuring effective collaborations. As a preliminary analysis, we found that various indicators captured by the Kinect sensor were correlated with participants’ quality of collaboration (e.g., amount of talking and movements). Finally, we are showing that well-designed learning activities can teach beginning computational thinking skills to a variety of participants - even those with no programming experience.

In the future, in addition to developing more multi-modal measures of collaboration, we are planning to improve the activity by making it longer and by providing more scaffolding. We will also design alternative ways of displaying the amount of talking, for example by making it more salient in the environment. Finally, we are interested in studying longer-term activities, for instance when students are working together over the span...
of several days or weeks. This will address an important limitation of most studies where collaborative episodes take place in a short time frame (i.e., 1-2 hours).

References


Social and Cognitive Group Awareness to Aid Argumentation About Socially Acute Questions on Social Media

Marielle Dado, University of Duisburg-Essen, marielle.dado@uni-due.de
Daniel Bodemer, University of Duisburg-Essen, bodemer@uni-due.de

Abstract: Debates about socially acute questions (e.g., migration) may help develop argumentation skills. However, students may be hesitant to present different views to maintain interpersonal relationships, hindering communication and integration of multiple perspectives. Social networking sites (SNS) have been used to extend classroom discussions to such real-world topics. However, their flat-structured layout may not suit argumentation activities. This quasi-experiment in an applied classroom setting investigates the effects of a group awareness tool (GAT) that combines social (group members’ names) and cognitive (discussion points/stance in initial arguments) information in a network graph to aid communication behavior and integration of multiple perspectives during argumentation on a flat-structured SNS. Students supported by the GAT engaged in discussions with non-friends and students outside their class more than the control group, though the latter integrated multiple perspectives more. The GAT appears to have increased familiarity among non-friends. The potential influence of interpersonal relationships on integration of multiple perspectives is discussed.

Introduction
Interpersonal relationships have been found to influence learning outcomes, especially when it comes to discussions about socially acute questions (SAQs): social dilemmas that are often controversial and that have implications on several fields of knowledge (Simonneaux, 2007). Students tend to be exposed to media representations of SAQs that are socio-sociological (e.g., immigration, globalization) or socio-scientific (e.g., global warming, cloning) in nature, which means that students may already have personal opinions about these issues (Zeidler & Nichols, 2009). Thus, introducing SAQs in the classroom provides an opportunity for differing views to emerge, and therefore for argumentation skills to develop. However, given the controversy that often surrounds SAQs, students may be hesitant to express disagreement in the interest of maintaining harmonious interpersonal relationships (Kuhn, Wang, & Li, 2010). This is a challenge because demonstrating and acknowledging awareness of opinions and perspectives other than one’s own is a marker of good argumentation quality (Sadler & Donnelly, 2006). In order to develop argumentation skills, students should be exposed to multiple perspectives through social interaction and discourse in order for them to evaluate claims, analyze evidence, and make informed judgements about these issues (Simonneaux, 2007; Leitão, 2000).

Recently there has been growing interest in the use of social networking sites (SNS) such as Facebook and Google Community to support discussion and mutual learning between students (Manca & Ranieri, 2013). One big motivation for teachers to adopt SNS in their classroom is to have a platform to discuss real-world issues related to the current lesson (Chen & Bryer, 2012). This suggests that social media may be an appropriate platform to discuss SAQs. However, since many SNS platforms were built for commercial rather than learning purposes, they may lack structural features that promote meaningful discussions between learners. Kirschner (2015) particularly criticized the “flat structure” arrangement of discussions on SNS: that is, unlike hierarchical and threaded discussions, posts on a typical SNS appear all in one page in reverse chronological order, with replies to each post appearing un-nested below each post. This makes it difficult to find one’s own postings, let alone those of others (Kirschner, 2015). In an argumentation activity, this could make it difficult for learners to become aware of others’ discussion points and to connect their claims to the arguments of their peers. To overcome this limitation, group awareness tools (GATs) could be designed to provide information through visualizations about members of a learning group to implicitly aid individual learners to behave, communicate, and reflect in ways that are productive for collaboration and learning (Janssen & Bodemer, 2013). This information can be broadly categorized as (1) social/behavioral awareness, such as who are the group members, who they are communicating with, or how they are contributing to the task; or (2) cognitive/knowledge awareness, such as the level of prior knowledge of group members, the information that they possess or opinions that they hold. When GATs are designed to allow for comparison of one’s knowledge or behavior to that of their peers’, learners may adapt their actions accordingly. For example, GATs that highlight knowledge differences led learners to communicate with their peers to fill their knowledge gaps (Erkens, Schlottbom, & Bodemer, 2016) or to discuss perceived conflicts with peers in a more interactive way (Bodemer, 2011).
Studies about GA support specifically for social and attitudinal characteristics of group members during argumentation in SNS have yielded some promising results. In a series of controlled lab experiments, Tsovaltzi and colleagues supported group awareness on a Facebook-like platform by informing participants, as they were preparing their individual arguments prior to collaborative discussion, that their arguments may be published after completion of the experiment for other students to comment and amend. In one study (Tsovaltzi, Puhl, Judele, & Weinberger, 2014), they found beneficial interaction effects on individual argument elaboration when combined with argumentation scripts, but a detrimental main effect on learning. In another study, they additionally found that group awareness did not lead to considering multiple perspectives (Tsovaltzi, Judele, Puhl, & Weinberger, 2015). The researchers hypothesize that group awareness in this case led to overcautiousness whilst constructing their initial arguments, which led to more individualistic behavior. Puhl, Tsovaltzi & Weinberger (2015) found that displaying in a 2-dimensional space the communication attitudes of the group members’ relative to one’s own attitudes led to gains in domain knowledge and a change of attitude towards multi-perspective communication.

Previous work suggests that the interpersonal relationships between students may have an influence on their awareness of multiple perspectives, which is essential for successful argumentation. Although previous GAT research in argumentation in SNS have demonstrated some promising effects, none so far have addressed overcoming the “flat-structuredness” of SNS by bringing awareness to the diversity of arguments in the community and enabling comparison of one’s opinions to the opinions of others. The present study investigates the effects of a GAT that attempts to address these concerns: by combining social information (names of group members) and cognitive information (discussion points and stance of each group member) in a network graph to aid argumentation on a flat-structured SNS. Particularly, the influence of GA support on communication behaviour, learning outcomes, shifts in opinions, awareness and integration of multiple perspectives are evaluated in a SNS in applied classroom setting, which is different from most GA studies on argumentation in SNS which are usually conducted in laboratory settings (Tsovaltzi et al, 2015; Bodemer, 2011). In addition, the study also describes how users interact with a GAT arranged in a network graph and what these interactions imply about how the information contained in it is processed productively.

Method

Design and participants
A quasi-experiment was conducted with twenty-nine Year 12 students (mean age=17.2, SD=.45) in two Economics classes with the same teacher in an International Baccalaureate (IB) school in Germany. Because the experiment sessions were embedded as classroom activities, the research design was selected in the interest of ecological validity and in order not disrupt the remaining class time. One class (n=14) was randomly assigned as the control group and the other (n=15) was assigned as the GAT (i.e., experimental) group. Their final Economics grades in Year 11 (7 is the highest possible grade) shows that prior academic performance between the control (M=6.2, SD=1.17) and GAT (M=5.86, SD=1.1) groups is similar. In terms of English language ability, most of the students are placed in English A class (native/native-like speakers), except for 1 student in the control group and 3 students in the GAT group. However, the teacher believes that both classes are competent in academic English, and the English B students perform well in their economics classes. Six Year 12 students (all 17 years old) from an IB school in Australia also took part in the experiment as a classroom activity and interacted with participants; however, due to technical difficulties their data was excluded from the analysis.

Social media learning environment, learning activity and instructions
The social media platform used for the study is called Google+, an SNS created by Google, which allows users to create “Communities” in which users can engage in asynchronous discussions about specific topics. The format can be described as flat-structured (e.g., posts appear chronologically and replies to posts were not nested). This platform was selected by the IB Economics teacher as it could be easily integrated into the students’ existing web tools, since all student email addresses were hosted by Google’s email service Gmail, which is required to access Google+. The teacher created a private Google Community for his 2 Economics classes, as well as the economics class of a former colleague in Australia, to provide a platform for debates about real-world issues related to economics. Apart from a brief session in which each student had to post a short self-introduction, participants had no prior experience with Google+ before the experiment sessions.

Four experiment sessions were integrated as class activities in 4 separate 50-minute Economics class periods over the course of 4 weeks (1 session per week). Both groups completed Sessions 1, 3 and 4 on the same days. Due to some technical problems, Session 2 for the GAT group was postponed to the day after the control group completed Session 2. No class was allowed to proceed to the next session until all students have
participated in the current one. Students worked on the experiment sessions on their own computers and were not allowed to engage in off-task behavior or speak to their classmates. They were also not allowed to interact on Google Community outside experiment sessions.

The first 3 experiment sessions corresponded to the knowledge building cycle proposed by Leitão (2000) whereby students first (1) construct an initial argument (Session 1); (2) have the opportunity to construct a response/counterargument (Session 2); and (3) construct a reply to responses they received and revise initial argument (Session 3). “Migration” was selected as the topic for discussion for its relevance to the Development Economics unit of the IB Economics syllabus and informal observations that suggested a diversity of perspectives among the students. In Session 1, each student was asked to create a post on Google Community with a 4-6 sentence argument to the question: “To what extent do you agree with the statement: ‘Migration from developing to developed countries leads to economic development.’?” Each post had to (1) mention at least 1 economic stakeholder and whether it benefits from migration; and (2) had to be supported by at least 1 article from a credible online resource (students were provided with links to appropriate sources). In Session 2, both control and GAT groups were given the same instructions to follow when commenting, namely (1) comment on at least 1 post; (2) structure comments around their initial arguments in Session 1; (3) if they hold the same perspective, write down exactly what they agree on and why; (4) if they hold a different perspective, write down exactly what they disagree on and why; (5) ask questions if something is not clear or if they would like more information; (6) give feedback or suggestions and explain why they think this could be helpful. Students were also explicitly instructed to read as many posts as they can before commenting, as well as how to use the “Search” function to find posts by specific people, so that they do not simply comment on the first post they see. In Session 3, students were instructed to (1) add to or edit their Session 1 argument if their perspective has changed, or to write new 4-6 sentence response to the initial question; or (2) write about why their perspective has stayed the same. In Session 4, students answered a survey about their experience during the experiment.

Group awareness tool

The GAT provided to the GAT group visualizes how each student’s initial argument (Session 1) is positioned with respect to the opinion/responses of their peers (see Figure 1). The tool is a network graph composed of 2 types of nodes: (1) students (circles); and (2) migration stakeholders (rectangles). An edge is drawn between these 2 node types if the student mentions that stakeholder in their initial argument, with the color (green or red) indicating the stance of the student (whether it benefits or does not benefit from migration, respectively). Two coders jointly identified 11 stakeholders from the initial arguments and independently coded the stances (Cohens’ $k=0.86$). The network graph was then created using the R package visNet that produced an interactive graphic that allowed users to click and rearrange the graph elements. The following clickstream data was logged by the tool: drag view (i.e., zooming, panning and scaling the view of the graph), click node, drag node, click edge, drag edge. The tool itself is hosted on a separate URL and was not integrated on Google Community. Therefore, students had to keep two browser windows open during the experiment sessions (one for the tool, one for Google Community). The tool was intended for students in the GAT group to help them select which student (posts) to reply to in Session 2, based on their initial arguments on the discussion topic in Session 1. Somewhat similar to the GAT of Puhl et al (2015) on a 2D space, the networked arrangement of students and stakeholders (i.e., discussion points) allows not only a general overview of the initial arguments in the entire community, but also enables individuals to compare their arguments to those of others based on their connections (or lack thereof) to the same stakeholders. Based on previous GAT research that also enabled users to compare their information to others (e.g., Erkens et al, 2016), this arrangement along with the color-coded edges should help students to easily identify differences between stances on the same stakeholders and use it as a point of discussion. Thus, in order to overcome the flat-structured layout of Google Community, the GAT was designed to influence communication behavior on the basis of social (names of group members) and cognitive information (discussion points and stance of each group member), specifically with regards to differences in opinion.
Procedure

In Session 1, students were first given 5 minutes to answer a survey that asked for their opinion on specific migration stakeholders (see “Measures…” section). Then they were asked to create individual posts on Google Community with their initial arguments regarding the topic of migration, as stated in the “Social media learning environment…” section. Students were also encouraged, but not required, to use the sentence starter “I agree/do not agree with the statement…” Students were given 15 minutes to post their initial arguments. In Session 2, the control group proceeded to read and comment on their peers’ posts on Google Community for 15 minutes, whereas the GAT group was first introduced to the GAT in two phases. First, students were trained on how to interpret the GAT using a small-scale version with dummy data and were prompted to identify student pairs that had the same/different opinion on the same stakeholder. Second, they viewed the GAT and were instructed to find their name and compare their Session 1 arguments with the arguments of their peers. They were explicitly told that the GAT is interactive, and that they could click and drag the graph elements. Students were then asked in a survey to identify at least 1 student that shared/did not share their opinion on the same stakeholder. They were given 5 minutes for the training phase and 10 minutes to explore the GAT, before being given a further 15 minutes to comment on at least 1 of their peers’ posts on Google Community, just like the control group (the tool was made available to them during this time). In Session 3, students were instructed to read their Session 1 arguments and reflect on whether their perspective has changed since reading, commenting and receiving comments on their posts. Then, in the form of a comment on their Session 1 post, they were instructed to revise their initial opinions and respond to the comments they received. They were given 15 minutes to complete Session 3. Session 4 was a post-study questionnaire that asked students about their experiences in the Google Community activity (see “Measures…” section).

Measures, research question, and hypotheses

Communication behavior is defined as how students chose to reply to certain posts over others on Google Community in Session 2. The initial opinions of each student as expressed in Session 1 (i.e., stakeholders mentioned and stance) were compared to the initial opinion of the student that they replied to in Session 2, and categorized as having: (1) the same opinion on at least 1 stakeholder; (2) a different opinion on at least 1 stakeholder; or (3) having no shared opinion on a stakeholder. Students were also asked to indicate who among the students in the Google Community they would consider as their friends to check whether friendships had any influence on their communication behavior. Integration of multiple perspectives was based on Leitão’s (2000) description of the impact of others’ perspectives from the discussion on one’s initial argument: multiple perspectives can be dismissed, localized (alternative perspectives are acknowledged but the initial argument is retained), integrated with initial opinions (e.g., allowing for some exceptions or conditions), or fully accepted. The final arguments were coded accordingly by two coders (kappa=.76).

To measure learning outcomes, the teacher administered two practice exams for IB Economics (as is standard practice in the school) before Session 1 and after Session 4. These covered topics from the Development Economics unit of the IB Economics syllabus. One question (worth 8 points) required students to apply their economics knowledge to evaluate a reference news article (pre-test: the impact of the involvement of China in Ethiopia’s economy; post-test: strategies that the Haitian government can use to improve their economy). As stated in the grading rubric, to successfully answer the essay question “students must offer a considered and balanced review that includes a range of arguments, factors or hypotheses. Opinions or conclusions should be presented clearly and supported by appropriate evidence”. Therefore, students were required to apply the same argumentation skills in both the exam question and the experiment sessions.

Another measure administered before and after the experiment was opinions on migration stakeholders. In Session 1 (before posting initial arguments) and Session 4, students rated 8 economic stakeholders of migration in both the host (developed) country and country of origin (developing) according to whether migration negatively or positively benefits them (1 – very negative impact to 5 – very positive impact). Awareness of the multiple perspectives was evaluated in Session 4 by asking students, for each of the 11 stakeholders appearing on the GAT, whether they perceived the overall opinion in the community of the impact of that stakeholder on migration as mostly positive, negative, or rather well-distributed between the two stances. One point was given for each correct answer for a total of 11 points. In addition to the dependent variables and the clickstream data logged by the GAT, self-reports on the experiences of the GAT group with the GAT were collected in Session 4. Students were asked to rate (1 – strongly disagree to 5 – strongly agree) 4 statements pertaining to the usefulness of the GAT in (1) representing their initial argument; (2) providing them with an overview of the Community’s opinion on migration stakeholders (3) helping students identify which of their peers agreed/disagreed with their opinions; and helping students decide which posts to (3) read and (4) comment on in Session 2. Participants were also asked to assess their own change in perspective: whether their
perspectives (1) completely changed, (2) did not change at all, (3) changed by integrating new perspectives with original ones, or (4) did not change, but they were able to acknowledge alternative perspective (i.e., localized, Leitão, 2000). They were also asked to specify the most influential source of their change in perspective.

Finally, several control variables were also analyzed to investigate how they may impact the findings on the dependent variables. First, argumentation quality of initial arguments in Session 1 were assessed based on whether a credible source was cited, as well as on Sadler & Donnelly’s (2007) rubric on “position and rationale” for argumentation, specifically whether or not the claims were grounded (“offers a coherent, logically consistent argument that includes an explanation and rationale for his/her position”, p. 1474). Second, argumentation moves in Session 2 responses were categorized as either as an (1) agreement; (2) disagreement (including local agreement) or (3) non-argumentative move (e.g., questions, off-topic comments). Finally, for each student, their initial argument in Session 1 and their reply to in Session 2 were compared to check whether students used their initial arguments as a basis for their peer discussions: (1) complete consistency; (2) partial consistency; and (3) not consistent at all. These variables were coded by two coders; Cohen’s kappa is .88 for argumentation quality, .87 for argumentation moves and .94 consistency of comments to initial arguments.

The overall research question for the present study is: Can a GAT depicting (1) students (social information) (2) discussion topics and (3) opinion stances (cognitive information) foster argumentation and learning on a social media platform? It is hypothesized that students supported by the GAT will post arguments on posts that reflect a different opinion than their initial arguments (communication behavior) (H1); be more likely to demonstrate integration (H2) of multiple perspectives; will receive higher scores on learning outcomes (H3); (2) and exhibit a greater shift in opinions on migration stakeholders (H4), and demonstrate better awareness (H5) of multiple perspectives.

Results

Table 1: Sample size per experiment session and final sample size associated with each variable

<table>
<thead>
<tr>
<th>Class size</th>
<th>Control</th>
<th>GAT</th>
<th>Total</th>
<th>Variables analyzed with this sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 1</td>
<td>13</td>
<td>14</td>
<td>27</td>
<td>Argument quality</td>
</tr>
<tr>
<td>Session 2</td>
<td>13</td>
<td>13</td>
<td>26</td>
<td>Communication behavior; argumentation moves; consistency of comments to initial argument</td>
</tr>
<tr>
<td>Session 3 and Session 4</td>
<td>12</td>
<td>13</td>
<td>25</td>
<td>Number of friends; learning outcomes, opinion shifts; (1) awareness and (2) integration of multiple perspectives; self-reported perspective change</td>
</tr>
</tbody>
</table>

Some students were unable to complete all experiment sessions due to absences. Thus, Table 1 summarizes the final sample sizes for each of the measured variables. The 6 Australian students participated in Sessions 1-4. However, in Session 2 they mostly interacted with their fellow classmates, except for 2 students who replied to 1 student in each group. Therefore, their influence on the data is expected to be minimal.

Descriptive information on Session 1 (initial arguments) and Session 2 (comments)

A total of 33 initial arguments were posted on Google Community in Session 1 (27 from the experimental groups and 6 students in Australia). The quality of arguments was quite high among most students: both groups were able to make well-grounded claims in their initial arguments ($p=1.00$, Fisher Exact Test). On average, students in the control group mentioned 2.7($SD=.99$) out of 11 stakeholders, of which they took a “benefits from migration” stance for 1.7 ($SD=1.49$) stakeholders and a “does not benefit” stance for 1($SD=.88$) stakeholder. The GAT group mentioned 2.23 ($SD=1.59$) stakeholders, of which they took a “benefits from migration” stance for 1.69 ($SD=1.37$) stakeholders and a “does not benefit” stance for .54 ($SD=.78$) stakeholder. The mean difference of the two groups on their “does not benefit from migration” stance on stakeholders was significant (Mann Whitney $U=48.5, p=.029, r=-.419$), indicating that the control group was more likely to express this stance in their initial arguments than the GAT group. In Session 2, control group received on average 1.08($SD=1.24$) comments and the GAT group received 1.15($SD=9$) comments. However, not all students were able to receive comments on their posts: 5 from the control group and 3 from the GAT group. In terms of argumentation moves, the proportion of students that wrote counterarguments in their comments was 0.62 in the control group and 0.38 in the GAT group, although this difference was not significant ($p=.546$, Fisher Exact Test).
if they were not supported by the GAT than if they were (the odds of students integrating multiple perspectives from the discussion in their final answers was 6.67 higher of the 4 categories in Leitão (2000): integration and dismissal. A chi-square test of independence revealed that 

$$M =$$

little information and that it was difficult to remember which students agreed or disagreed with them. The remaining 5 students mention that they did not decide which posts to comment on. One student noted that the graph helped her find posts that talked about 

Community” (and “The graphic helped me gain a general overview of the perspectives of other members in Google 

the statements “I was able to see which members agreed or disagreed with my perspective” (control=4; GAT=6) that reading other students’ posts was the most influential in their change of opinions in the final answer; followed by comments received on their posts (mentioned by 3 students in each group).

Result: Both groups were also equally likely to use their initial arguments as the basis for the comments ($\chi^2(1)=.653, p=.419$).

**Dependent variables**

In terms of the learning outcomes, mean pre and post-test scores for the control group are 4 ($SD=1.1$) and 3.83 ($SD=1.6$); for the GAT groups, mean pre and post-test scores are 3.31 ($SD=1.3$) and 4 ($SD=1.6$), respectively. However, a mixed ANOVA yielded no significant interaction effect between time and the presence/absence of the GAT ($F(1, 23)=1.626, p=.206$). Analysis of communication behavior in Session 2 also did not yield any significant group differences ($p=0.667$, Fischer’s Exact Test). Both groups were equally likely to comment on peers’ posts that expressed the same opinion as theirs (and had no differing opinions) on the same stakeholder (control=6, GAT=8). However, students in the control group were 12.6 times more likely to comment on friends’ posts than students in the GAT group (95% CI: 1.19, 678.9, $p=0.03$, Fisher Exact Test). Furthermore, the odds that students would comment on posts written by their classmates were also 5.33 higher if they belonged to the control group ($\chi^2(1)=4.06, p=.044$). This finding is interesting considering that both groups tended to have friends from within and outside their classroom. The average numbers of friends in the same class were 3.5 ($SD=1.9$) for the control group and 4.46 ($SD=2.9$) for the GAT group, though this difference was not significant (Mann Whitney $U=69.5, p=.63$). In terms of friends in the other Economics class, the average numbers are 2.25 ($SD=1.86$) and 3.85 ($SD=2.82$) for the control and GAT groups, respectively, although the difference is non-significant as well (Mann Whitney $U=48.5, p=.104$). A factorial ANOVA with aligned ranks transformation yielded no significant interaction effect between time (pre-test/post-test) and the presence/absence of the GAT in terms of shifts of opinions regarding the 16 rated stakeholders ($p<.05$). In terms of awareness of multiple perspectives, the mean score for the GAT group was 5.8 ($SD=1.1$) and the control group mean score was 5.2 ($SD=1.6$), although the difference is non-significant (Mann Whitney $U=53.5, p=.171$). Finally, for integration of multiple perspectives, the final arguments of the students fell into only 2 out of the 4 categories in Leitão (2000): integration and dismissal. A chi-square test of independence revealed that the odds of students integrating multiple perspectives from the discussion in their final answers was 6.67 higher if they were not supported by the GAT than if they were ($\chi^2(1)=4.98, p=.026$). However, according to self-reported change in perspective, only 2 students in each group reported that their perspective “did not change at all”. The self-reports also indicate that of the 21 students who reporting integrating new perspectives, 10 of them (control=4; GAT=6) that reading other students’ posts was the most influential in their change of opinions in the final answer; followed by comments received on their posts (mentioned by 3 students in each group).

**How the GAT group made use of and perceived the group awareness tool**

In Session 2, the GAT group spent an average of 3.72 minutes ($SD=1.9$) exploring the GAT. There were 240 recorded interactions, most of which (84.6%) were of students zooming into, panning and scaling the view of the graph. This means that it is not possible to determine from the log data which elements of the graph students were paying attention to, or whether they noticed similarities and differences between student opinions. However, in the training phase in Session 2, all 14 students were able to correctly identify student pairs from the dummy data that had the same/different opinion on the same stakeholder. All 14 students were likewise able to correctly identify from the GAT at least 1 student each that shared/did not share their opinion on the same stakeholder, although only 5 students actually posted on the posts by the students they identified. In Session 4, the GAT group gave neutral to positive ratings on the usefulness of the tool. Students gave positive ratings to the statements “I was able to see which members agreed or disagreed with my perspective” ($M=4.38, SD=.77$) and “The graphic helped me gain a general overview of the perspectives of other members in Google Community” ($M=4.3, SD=.85$). However, they gave neutral ratings when asked about how the tool represented their perspective ($M=3.62, SD=.76$) and whether the tool helped them decide which posts to read ($M=3, SD=1.2$) and comment on ($M=3.30, SD=1.38$) in Session 2. When prompted to explain their ratings, 7 students mentioned that seeing who shared or did not share their opinions on the same stakeholder helped them decide which posts to comment on. One student noted that the graph helped her find posts that talked about stakeholders that she did not mention in her initial argument. The remaining 5 students mention that they did not take the tool in consideration when deciding on posts to comment on, stating that it was confusing, provides little information and that it was difficult to remember which students agreed or disagreed with them.

**Discussion**

The present study investigated a GAT for argumentation about a socially acute question on a flat-structured SNS, which is often criticized as not being conducive for argumentation activities. Like previous GAT designs for argumentation in SNS, it attempted to provide both an overview of group information and encourage
comparison of one’s information to others. In this study, however, the GAT combines two kinds of GA information (social and cognitive) and was designed to guide a specific phase in the activity, whereby students must select peer group members to discuss their opinions on migration with. This was an important consideration given that prior to the experiment, the participants already had relationships with each other, as well as opinions about migration, which could influence how they behave during argumentation activities.

In terms of communication behavior, no significant differences were found on the dependent variable measured (H1): both groups mainly conversed with students that expressed the same initial arguments as theirs on at least one stakeholder. It should be noted that both groups were equally likely to express approving and dissenting opinions during discussion. This could imply that discussions among students probably went beyond the topics (stakeholders) that they had in common. Based on the log files, interactivity with the details (e.g., nodes and edges) of the GAT was quite low, even though students were explicitly told that they could interact with the graph. Thus, it is likely that the GAT group did not deeply process the commonalities and differences between them and their peers. However, subjective ratings indicate that students supported by the GAT may have noticed the differences in perspectives in the group. Furthermore, students were able to correctly interpret network graphs with dummy and real data; thus, the networked visualization itself was not too complex to comprehend. However, from a technical perspective, clicking a name or stakeholder (nodes) as well as their stance (edges) did not automatically redirect users to the corresponding posts on Google Community, which could have discouraged the GAT group from interacting in Session 2 in accordance to the GAT. Nevertheless, despite having friends in both classes, students in the control group overwhelmingly decided to respond to their friends in the same class, whereas students in the GAT group were equally likely to comment on friends’ and non-friends’ posts. The control group’s behavior is consistent with studies showing that people tend to be more willing to express their opinions online to friends (Luarn & Hsieh, 2014). Studies have shown that unfamiliarity with community members demotivates active participation in online discussions (Preece, Nonnecke, & Andrews, 2004) and that familiarity increases the likelihood of participation in online discussion forums (Hew, Cheung, & Ng, 2010). Given that the most dominant action captured by the GAT is zooming, panning, and scaling implies that the GAT group paid more attention to the GAT as a whole, without dwelling too long on any finer node-edge relations. It is possible, then, that they were able to get an overview of who is in the community and which stakeholders they had an opinion about. Hence, the GAT may have created a sense of familiarity among the students and help them consider the posts of peers that they are not personally close to.

Another finding of the study is that the control group was more likely to integrate multiple perspectives from the discussion than the GAT group (H2). Since not all students in the control group received comments on Session 2, the integration of multiple perspectives could not have only been the result of receiving comments containing a different perspective; as suggested by the self-reports, reading posts with dissenting opinions may have also had an influence. However, the study was not able to track which posts were read by whom and whether the GAT influenced this behavior. Nevertheless, other studies have suggested that opinion change is influenced by the relationship of a person with their communication partners. People are more inclined to change their opinion when exposed to the opinions of people who are closely related to them; specifically, any dissimilarity in initial opinions is reduced by social influence (Friedkin & Johnsen, 1997). Even when friends tend to have the same initial opinions, friends are still more likely to express disagreement anyway, which could foster learning from opposing perspectives (Morey & Hutchens, 2012). This could explain why the control group, who mainly commented on each other’s posts, were more open to changing their initial opinions after being exposed to the opinions of friends. When people are not closely related, as in the case of the GAT group students, then the social influence process does not necessarily influence any changes in initial opinions (Friedkin & Johnsen, 1997). Another explanation could be that there were more diverse opinions on migration among the control group’s initial arguments, as they were found to be more likely to take a “does not benefit” stance than the GAT group. Thus, the control group could simply have been more open-minded and receptive to different views after discussion (Barabas, 2004). Furthermore, self-reports show that only 4 students (2 in each group) indicated that their perspective did not change at all. It is perhaps possible that more students in both groups were able to integrate or localize new perspectives from the discussion, but did not express this when they revised their opinions in Session 3.

Overall, the study demonstrates that combining social and cognitive information in a GAT with a network visualization may influence communication behavior, but not necessarily in accordance to awareness of differences in initial arguments. Rather, the GAT helped students find posts in a flat-structured SNS that were written by students with whom they do not have a close personal relationship, potentially increasing familiarity among students in the community. Building familiarity could be an important consideration for argumentation in SNS to encourage students to freely discuss controversial SAQs in a supportive environment. The results further suggest that communicating with friends and exposure to the opinions with whom one has a close
personal relationship may influence the likelihood of adopting new perspectives, although alternative explanations cannot be ruled out (i.e., that the control group could have been more open-minded). Thus, the seemingly contradictory results (i.e., familiarity may increase overall engagement among non-friends, but communication among friends may lead to more multiple perspective taking) could be further investigated in relation to other variables (e.g., open-mindedness). As these issues will continue to be relevant inside and outside the classroom, it would be beneficial to understand how to best encourage students to discuss socially acute issues in a meaningful way.

References


Learning Analytics in Support of Qualitative Analysis

Bruce Sherin, Northwestern University, bsherin@northwestern.edu
Nicole B. Kersting, University of Arizona, nickik@email.arizona.edu
Matthew Berland, University of Wisconsin–Madison, mberland@wisc.edu

Abstract: The purpose of this manuscript is to describe a style of learning analytics research that makes closer contact with the qualitative methods of the learning sciences. We do so by drawing on two examples of existing research. The first is our prior work attempting to automate part of the Classroom Video Analysis assessment, which seeks to measure knowledge of teachers that is predictive of their performance. The second is from work which looks at how middle school students explain the Earth’s seasons. In that work, we attempted to use unsupervised methods to capture elements of what was previously a fully qualitative analysis. Our goal is to provide the reader with a sense for this style of research that brings qualitative analysis and analytic methods into closer contact. To accomplish this, we make use of a new system called Tactic, designed specifically to support this mode of research.

Introduction

The purpose of this manuscript is to argue for a style of learning analytics research that makes closer contact with the qualitative methods of the learning sciences. As a set of innovations in educational research, learning analytics is not only about the application of new analytic algorithms to data. These new algorithms are also applied to new types of data, such as computer log files and biometric data. They are also applied to data that is larger in scope. Furthermore, the types of questions being answered are often somewhat different (Baker & Yacef, 2009; Martin & Sherin, 2013).

However, it is possible to apply some techniques from learning analytics to more traditional qualitative data, and with aims and methods that hew more closely to established methods. Why would we want to do this? There are a number of obvious reasons. First, if we apply traditional and learning analytic methods to similar corpora and with similar aims as existing studies, we have the potential to produce distinct but complementary analyses. A second benefit is the potential for reducing the labor associated with traditional types of analysis, especially in qualitative research. Even when a data corpus is small, from the point of view of learning analytics, manual analysis can be slow and laborious.

But, more exciting than these somewhat obvious benefits is the possibility of bringing established and learning analytic methods together in a manner that augments both. If we work with smaller data corpora about which human analysts have detailed, intimate knowledge, there is the potential to support the design of more finely tuned computational analyses, and to better interpret the results of those analyses. Computational analysis can in turn provide new perspectives on qualitative data analysis, even helping to surface tacit knowledge of human coders. Finally, tightly integrated traditional and computational analyses have the potential to provide a kind of triangulation that increases our understanding of—and confidence in—both.

Our purpose in this paper is to illustrate these points. We do so by drawing on two examples of existing research. The first is our prior work attempting to automate part of the Classroom Video Analysis assessment, an assessment that seeks to measure knowledge of teachers that is predictive of their performance (Kersting, 2008; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Kersting, Givvin, Sotelo, & Stigler, 2010). In that analysis, we used data coded by human raters to train classifiers using a modified Naïve Bayes approach (Kersting, Sherin, & Stigler, 2014). The second is from work which looks at how middle school students understand the Earth’s seasons (Sherin, Krakowski, & Lee, 2012). In that work, we used unsupervised methods to capture elements of what was previously a fully qualitative analysis (Sherin, 2013). Both of these lines of work followed a similar trajectory; in each case, we revisited prior research, applying learning analytic methods to existing data.

Tactic: Infrastructure for learning analytics research

We begin with a brief discussion of the infrastructure available for this new style of learning analytics research. Learning analytics requires, of course, software tools that can perform the relevant computational analyses. One approach that analysts can take is to write their own code, usually in a way that is specific to the analysis at hand. Even in this case, however, analysts draw on publically available software libraries. In addition, there do exist GUI tools that reduce—but generally don’t eliminate—the programming required. These include Weka (Hall et al., 2009), RapidMiner (Hall et al., 2009), and LightSide (Mayfield & Rosé, 2013).
In this paper, we will display analyses as they look in a new web-based environment text mining environment called Tactic, that we are currently developing. Tactic is specifically tuned for working in an interactive mode in which the analyst moves back and forth between the data and analytics. The data is always front and center, and Tactic incorporates visual features that assist the user in seeing relationships between the original data and analytics applied to that data. In addition, Tactic does not impose a particular decomposition of the analytic process on users. So, for example, we do not assume that a user’s analysis will be separated into a pre-processing step, followed by a modeling step. We believe instead that the interactive style of research should allow for a more emergent workflow, which might be somewhat specific to each project.

Figure 1 shows an analysis workspace, in process, with two tiles visible on the right. (This will be further explained below.) Tiles do most of the computational work, and Tactic includes a base set of tiles which users modify or use unchanged. Users can also create entirely new tiles. All programming is done in Python, within the web environment. There is nothing in Tactic that represents a radical advancement beyond existing tools. What is important is that it is tuned for a particular style of work, which we now explain in more detail.

**First example: The CVA assessment**

**Original CVA research**

The second author of this manuscript and her colleagues developed the Classroom Video Analysis (CVA) assessment as a means of assessing the usable knowledge of mathematics teachers (Kersting, 2008; Kersting et al., 2012; Kersting et al., 2010). Here, by “usable” knowledge, we mean the knowledge that teachers can access and apply in the classroom. In the CVA, teachers watch a series of videos of classroom instruction. After each video, they are asked to “analyze how the teacher and student(s) interacted around the mathematical content.” They type their comments into a web-based form, and these responses are then scored by trained coders. For example, in response to a video clip showing part of a lesson on fractions, a teacher responded: “The teacher is asking the student questions to narrow down his search to find equivalent fractions. I don’t know if the boy understood the meaning of the giant 1. It was almost like the student was guessing, and judging how the teacher responded he would know if he was right or wrong.”

The CVA has been the focus of an extended program of traditional (non-computational) research. That work has found that the CVA can predict instructional quality, as measured by direct observations of teaching. Strikingly, one dimension measured by the CVA, teachers’ ability to make suggestions for improving the teaching episode, has been found to predict the learning of students in the classrooms of teachers studied.

**Prior computational methods**

The CVA is potentially a powerful means of assessing teachers. However, its wider use has been somewhat hindered by the fact that it is laborious for coders to score teachers’ written responses to the video clips and requires coders to undergo extensive training. For this reason, automating some part of the CVA analysis could have substantial benefits. In addition, the power and success of the CVA merits additional study. Learning analytic methods have the potential to not only replicate the work of manual coders; they might also help us to better understand why and how coders assign the codes that they do.

In prior computational work on the CVA (Kersting, Sherin, & Stigler, 2014), we made use of CVA assessments targeting three mathematical content areas: (1) fractions, (2) ratios and proportions, and (3) variables, expressions, and equations. For the fractions data, 238 teachers responded to 13 clips. For the ratio and proportion data, 238 teachers responded to 13 clips. Finally, for the variables, expressions, and equations data, 249 teachers responded to 14 clips.

All of this data was manually scored according to four rubrics: mathematical content (MC), student thinking (ST), suggestions for improvement (SI), and depth of interpretation (DI). Each response was given a code of 0, 1, or 2 for each of these four rubrics, with 2 being the best score, and 0 being the worst. For example, for the MC rubric (on which we will focus) a response was given a score of 0 if it didn’t mention mathematical content. In contrast, it was given a score of 2 if there was an in-depth analysis of the content, which went beyond the mathematics described in the clip. A more superficial discussion of the mathematical content received a score of 1. For instance, the teacher response cited above was given a score of 1 because it mentions the mathematical content but doesn’t go beyond the mathematics described in the clip.

In our computational analysis, we trained a unique Naïve Bayes classifier for each of the 40 clips. In essence, each classifier is a software model, that can take a teacher response as an input, and output the most likely code. The model is initially “trained,” by feeding in sample responses that have been coded by human analysts. In the case of the Naïve Bayes classifiers we constructed, this training is essentially done by gathering statistics; the algorithm looks at how frequently specific individual words are associated, in the human-coded
data, with a specific code. Furthermore, to train our classifiers, we did not include all of the words that appear in responses. We first discarded all words that appeared on a “stop list” of common words. We then kept only the 100 most frequent words that remained. After training, the Naïve Bayes algorithm combines these statistics to code new responses. More expert readers will recognize that, from the point of view of machine learning, the Naïve Bayes algorithm we have selected is relatively simple, and likely not optimal. The simple nature of our selected method was driven, in part, by our desire to produce classifiers with internals that could be easily inspected and understood, as we illustrate below.

In our computational work, one fundamental question was whether we could ultimately hope to replace human coders with automated scoring, as a means of predicting teacher quality. If this is our goal, it means that we should not be primarily interested in replicating individual codes on specific clips. Rather, we are interested in the ability of the automated coding to predict teacher effectiveness. This leads us to be most interested in (1) the composite scores, across all clips, for a given subject-matter area, and (2) the correlations between these composite scores and the composite scores given by human coders.

Table 1: Average quadratic weighted kappa and correlations for the CVA analysis

<table>
<thead>
<tr>
<th>Topic</th>
<th>Kappa</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractions</td>
<td>.51</td>
<td>.89</td>
</tr>
<tr>
<td>Ratio</td>
<td>.51</td>
<td>.86</td>
</tr>
<tr>
<td>VEE</td>
<td>.55</td>
<td>.91</td>
</tr>
</tbody>
</table>

Table 1 shows the average results for this analysis, in which the composite scores are averaged across multiple runs. As shown in the table, overall average correlations were high. The correlations for three of the four sub-scores – MC, ST, and DI – were similarly high, ranging from .77 to .91. For SI, the correlations were somewhat lower, ranging from .49 to .69. (The correlations for sub-scores are not shown in the table.)

We also computed quadratic weighted kappa values to provide a measure of agreement, beyond chance, between human and automated scoring. The results shown in Table 1 report the Kappa values for each of the subject matter areas, averaged across clips. These values represent moderate agreement. As with the correlation results, there was variation across the four rubrics. Results for MC and DI ranged from .56 to .64. In contrast, results for SI and ST ranged from .36 to .43 and .43 to .47 respectively.

Interactive analysis with tactic

The results summarized above suggest that applications of computational methods to the CVA data can achieve one of the benefits mentioned earlier; namely it seems to have the possibility of reducing the labor required by human analysts. We next want to show what the benefits might be for a more interactive style of analysis.

Figure 1 shows an analysis of the CVA fractions corpus, as it might appear in Tactic, at an intermediate stage. The data table is on the left. It has the text of each response. In addition, the column labeled CODEMC has the codes given by the human raters – 0, 1, or 2 – for the MC rubric. (Here, for simplicity, we will focus just on that rubric.) The popup list above the table allows the user to select other documents in the corpus, which here correspond each of the 13 fractions clips.

Analysis tiles are added using the menus at the top of the display. Once a tile has been added to the environment, the user clicks on the gear icon at the top-left of the tile to configure options on the back of the tile. Here three analysis tiles have been added to the work environment. The first, very basic, example, is the tile labeled WordFreqDist. When this tile is configured and run, it displays a table of the corpus frequency and document frequency for the most frequent words. The tile also provides a simple type of interactivity; namely, clicking on a word highlights matching words in the main data table. In addition, other tiles can access python data structures that are exported by a tile. In this case, there is a tile showing a plot of the document frequencies exported by the WordFreqDist tile.

The third tile allows us to illustrate some more interesting possibilities. This tile is set up to run a very simple Naïve Bayes analysis—a much more rudimentary analysis than was used in (Kersting, Sherin, & Stigler, 2014). Here we are training one classifier to work across all 13 of the fractions clips. In addition, (for the purposes of keeping things simple) we are committing what is usually an egregious error; we are training and testing on the entirety of the data.

When run, this tile trains a classifier using all of the fractions data. It then codes each response, and writes those codes into a new column, here labeled MC_AUTO. In addition, a confusion matrix is displayed on the front of the tile. Again, this provides some simple interactivity: clicking in a cell of the confusion matrix.
shows a new table of responses that correspond to that cell. (This isn’t shown in the figure.) Those responses can, in turn, be clicked to be viewed in context in the main data table.

Figure 1. Analysis of the CVA fractions corpus in process.

![Figure 1](image1.png)

However, a more interesting type of interactivity is possible, if we provide the qualitative data analyst with a window into the automated analysis. The Naïve Bayes tile shown in Figure 1 has three buttons, one corresponding to each of the three codes: 0, 1, and 2. Clicking on one of these buttons causes the text in the main table to be highlighted with varying colors, with the colors selected according to the conditional probability of the highlighted word, given the code corresponding to the clicked button. The color palette used goes from red to yellow to green. So, for example, dark green indicates a high probability of the word for the given code, and dark red a low probability.

Figure 1 shows the coloring obtained when the button corresponding to a score of 2 is pressed. Again, this is the highest code, and a score of 2 represents a deep analysis of mathematical content. Thus, when colored in this way, we are seeing which words the coders associate with a desirable response on this dimension. Figure 2 shows how the response we quoted earlier is colored. The word “fractions” is dark green, indicating a high probability. This is not surprising, given that the coders are looking at whether the teachers are attending to the mathematical content. In contrast, the words “right” and “wrong” are less indicative of a deep analysis. This is also perhaps not surprising – coders might well believe that saying that an answer is right or wrong implies a focus on superficial aspects of the mathematical content.

![Figure 2](image2.png)
Second example: Seasons analysis

Original seasons research
In this second example, we apply computational methods as part of what was originally a program of pure qualitative research (Sherin, Krakowski, & Lee, 2012). The work draws on interviews in which middle-school students were asked a range of questions about the Earth, its climate, and space science. More specifically, it draws on a portion of these interviews in which the students were asked to explain the Earth’s seasons.

The interviews were conducted in the manner of a clinical interview, which meant that the interviewer had the freedom to ask follow-up questions for clarification, and to further probe the student. The seasons portion of the interview began with the interviewer asking, “Why is it warmer in the summer and colder in the winter.” The interviewer then asked various follow-up questions. This included asking the student to draw a picture. Interviewers were also prepared with challenges to be introduced based on the initial explanations provided by students.

The explanations given by students were quite varied. However, in describing the spread of explanations, we have found it helpful to have in mind three prototype explanations. The first prototype we call close-farther explanations. In a closer-farther explanation, the whole Earth moves so that it is sometimes closer to the sun and sometimes farther away. When it is closer to the sun, the whole Earth experiences summer. When it is farther, the whole Earth experiences winter.

The second type we call side-based explanations. In a side-based explanation, the Earth moves in such a manner that one side, at times, faces toward the sun, while the other side faces away. (Usually—though not always—this is caused by the rotation of the Earth on its axis.) The side facing toward the sun experiences summer, and the side facing away experiences winter.

The final type of explanation is tilt-based. In these explanations, the Earth somehow moves so that one hemisphere is tilted toward the sun and the other away, with the side tilted toward the sun experiencing summer. This category includes, but is not limited to, the scientifically-accepted answer. In the accepted answer, the seasons are intimately linked to the Earth’s tilt. The Earth’s axis always points in the same direction. But, because the it orbits around the sun, first one hemisphere, then the other, will be inclined toward the sun. Furthermore, the hemisphere of the Earth tilted toward the sun receives more direct sunlight, and this more direct sunlight causes warmer conditions.

In the original work on this corpus, our argument was not that students each give one from amongst a small set of explanations. Rather we argued that students generally construct explanations out of a number of small components, which we called nodes, leading to a larger number of explanation structures that mix and match these components. We called these explanation structures dynamic mental constructs, or DMCs. Furthermore, we presented evidence, in the form of examples, that these DMCs could shift rapidly, over the course of an interview.

For illustration, we present a portion of the text of the interview with one student, who we refer to as Edgar. In his initial explanation, Edgar drew the diagram shown in Figure 3, as he said:

E: Here’s the Earth slanted. Here’s the axis. Here’s the North Pole, here's the south pole, and here’s our country. And the sun’s right here [draws circle on the left], and the rays are hitting directly right here, so things are getting hotter in the summer and when this thing turns, the country will be here and the sun can’t reach as much. It’s not as hot as the winter.

In our prior work, we argued that Edgar’s first explanation, given in the passage above, is a variant side-based explanation; the Earth spins, and the side facing the sun is warmer because it receives more direct sunlight. As the interview proceeded, Edgar seemed to recall that the Earth orbits the sun, in addition to rotating on its axes. This led him to transition to a fairly traditional closer-farther explanation.

Figure 3. Edgar’s seasons diagram.
E: Actually, I don’t think this moves [indicates Earth on drawing] it turns and it moves like that [gestures with a pencil to show an orbiting and spinning Earth] and it turns and that thing like is um further away once it orbit around the s- Earth- I mean the sun.

I: It’s further away?
E: Yeah, and somehow like that going further off and I think sun rays wouldn’t reach.

These brief excerpts illustrate, first, how a student explanation could shift, from one moment to the next, over the course of an interview. Edgar began with a side-based explanation, and shifted to a fundamentally different explanation, a closer-farther explanation. We can also get some sense for how explanations can be seen as built out of components, that are mixed and matched. For example, as part of his initial side-based explanation, Edgar makes use of the notion that more direct light causes parts of the Earth to be warmer, a component that we would more typically associate with a tilt-based explanation. However, here he has incorporated this idea as part of a side-based explanation.

Prior computational methods
The issues here are quite different than those encountered in the CVA research described above. In that case, there was a well-defined coding scheme, but that scheme was laborious to execute. Here, in contrast, the initial publications did not use a coding scheme at all; instead, they used examples from the corpus to illustrate and support theoretical claims. This meant that, in important ways, readers had to trust our interpretive ability – our ability to “see” the components and DMCs in the idea.

We believed that, if an automated analysis could replicate at least some aspects of our analysis, that would place the whole endeavor on more solid ground. Furthermore, we felt it was necessary to set the bar high. We did not want to specify, in advance, the components or DMCs that our automated analysis would use. The discovery of these elements is where the important, and controversial work, lies. Thus, we wanted to have the automated analysis discover, on its own, both the components and DMCs.

Going in, there were many reasons to think that the data might not be amenable to much in the way of a computational analysis. As we will see, the size of the data corpus is quite small. In addition, the interviews can be rambling and unclear. Furthermore, as in the interview with Edgar, there was a lot of gesturing, and references to drawings, important communicative elements that weren’t included in our computational analysis. These challenges are severe. But they are the sort of challenges that we will frequently need to overcome if we are going to use learning analytics in tandem with traditional qualitative data analysis.

Our computational analysis is presented in its most extended form in (Sherin, 2013). The data we used were the transcripts of interviews with 54 middle school students. As a first step in the analysis, all comments by the interviewer were removed, and utterances by the student in each interview were concatenated. The interviews were then broken into overlapping 100-word segments, using a moving window that moved forward in 25-word increments. So, the first segment had words 1-100, the next segment words 25-125, etc. This segmenting process resulted in a total of 794 segments, across all 54 of the interviews.

Words in a stop list of 782 words were removed from each of the segments. When this was done, the segments contained words from a vocabulary 647 unique words. Each of the 794 segments were then converted to word vectors, using a weight function of $1 + \log(tf)$, where $tf$ was the number of times the word appeared in the segment. What this means is that each segment of text was converted to a list of 647 numbers, where each number corresponded to one word in the vocabulary of 647 words. Finally, we performed one non-standard form of preprocessing. Namely, we computed what we have called deviation vectors. All of the 794 vectors were added, and the result normalized, to construct a sort of super-average. Then this average vector was subtracted from each of the 794 segment vectors. As explained in more detail in Sherin (2013), this was necessary for the next stage of the analysis to produce meaningful results. The result of the above process was that each segment of text was mapped to a point in a 647-dimensional space. As a final step, these points were clustered into 7 groups, using hierarchical agglomerative clustering. These clusters were interpreted as aligning with the knowledge components in our qualitative analysis. Ultimately, we argued that the analysis could in fact reproduce important aspects of the qualitative analysis, thus providing a new type of support for the original theoretical claims. We attempt to briefly illustrate this below.

Analysis in tactic
Figure 4 shows a Tactic workspace in which we have replicated some elements of the analysis reported in Sherin (2013). The top right tile in the workspace replicates the entire clustering analysis, described above. When this analysis is complete, the face of the tile displays a set of 7 tables, one for each of the clusters. These can be scrolled through (but only part of the first table is visible in the figure). To make the results more visible
here, these were sent to the Log area at the bottom of the workspace. These tables show the highest weighted words in each cluster.

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
<th>Cluster 6</th>
<th>Cluster 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>word</strong></td>
<td><strong>weight</strong></td>
<td><strong>word</strong></td>
<td><strong>weight</strong></td>
<td><strong>word</strong></td>
<td><strong>word</strong></td>
<td><strong>word</strong></td>
</tr>
<tr>
<td>moon</td>
<td>0.4494</td>
<td>closer</td>
<td>0.5245</td>
<td>spinning</td>
<td>0.3154</td>
<td>hemisphere</td>
</tr>
<tr>
<td>right</td>
<td>0.306</td>
<td>farther</td>
<td>0.4138</td>
<td>fall</td>
<td>0.5825</td>
<td>northern</td>
</tr>
<tr>
<td>day</td>
<td>0.2999</td>
<td>away</td>
<td>0.3225</td>
<td>spring</td>
<td>0.2448</td>
<td>sunlight</td>
</tr>
<tr>
<td>rotates</td>
<td>0.2867</td>
<td>point</td>
<td>0.2318</td>
<td>spins</td>
<td>0.2377</td>
<td>equator</td>
</tr>
<tr>
<td>earth</td>
<td>0.236</td>
<td>sun</td>
<td>0.1395</td>
<td>earth</td>
<td>0.2288</td>
<td>facing</td>
</tr>
<tr>
<td></td>
<td>0.118</td>
<td>axis</td>
<td>0.1395</td>
<td>axis</td>
<td>0.0806</td>
<td>north</td>
</tr>
<tr>
<td></td>
<td>0.087</td>
<td>cold</td>
<td>0.0742</td>
<td>guess</td>
<td>0.0958</td>
<td>hit</td>
</tr>
<tr>
<td></td>
<td>0.063</td>
<td>equator</td>
<td>0.057</td>
<td>seasons</td>
<td>0.5961</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td>0.021</td>
<td>warmer</td>
<td>0.0243</td>
<td>chicago</td>
<td>0.8966</td>
<td>towards</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>rotating</td>
<td>0.0192</td>
<td>time</td>
<td>0.0746</td>
<td>hit</td>
</tr>
</tbody>
</table>

Figure 4. Sample Tactic work environment for the seasons analysis.

The clusters are, we feel, relatively easy to align with the terms of our theoretical analysis. First, Cluster 2 has closer, farther, and away, as its top words. This suggests that the corresponding knowledge component could play a role in any explanation that focuses on the proximity of all or part of the Earth to the sun (i.e., a closer-farther explanation). There are also clusters we would most expect to see in side-based explanations; Cluster 7 has side as its highest-weighted word and rotates second. Cluster 3 talks about the spinning of the Earth. In addition, Cluster 1 is about the moon, day, and night. (Discussion of day and night often showed up with side-based explanation.) Finally, Clusters 4, 5, and 6 correspond to components we would typically associate with tilt-based explanations. Cluster 6 has tilted, toward, and tilt as its highest weighted terms, and Cluster 4 seems to be about the Earth’s hemispheres. Lastly, Cluster 5 appears to focus on the directness of light striking the Earth.

One of the core elements of our original work was the analysis of the dynamics of interviews. To capture this with our computational analysis, we performed a secondary analysis in which the clusters were applied back to each of the original transcript. The heatmap in Figure 4 shows the result when this analysis is performed for the full text of the interview with Edgar. To create this heatmap, the text of the interview was prepared and segmented just as for the clustering analysis, and a vector computed for each segment. The result, for the interview with Edgar, was 10 vectors. We then found the dot product of each of these vectors, with the centroids of each of the 7 clusters. In the heatmap, the time of the interview proceeds from left to right, and each
row of the heatmap corresponds to one of the 7 clusters. Darker shades represent a higher dot product between the section and cluster centroid. Thus, for example, the plot tells us that, the vector for segment 1 has its highest dot product with the directly-heat cluster, and segment 7 has the highest dot product with the closer-farther cluster. More generally, the plot does seem to broadly align with the qualitative analysis of the interview with Edgar. Recall that Edgar initially gave a side-based explanation, in which the side of the Earth facing the sun is warmer because it receives more direct sunlight. We see that, in fact, the segments in the first part of the interview do seem to align strongly with the directly-heat and side-rotate clusters. The first segment also seems to align with the tilted-towards cluster. This is because Edgar initially mentions the axis and poles. We also saw that, in the latter part of the interview, Edgar shifted to giving a closer-farther explanation. This is also captured in the heatmap. Thus, we can see that the automated analysis has captured at least some features of the dynamic account produced by the fully qualitative analysis, in our original work with this corpus.

Conclusion
Grounded theory, discovery-focused qualitative methods, and arduous qualitative coding have shown remarkable value for educational research over the last 50-100 years; but, so far, few environments or methods have leveraged learning analytics and machine learn to focus primarily on these analyses.

Our argument is that by using learning analytics as a tool for existing qualitative analyses, bringing the two fields closer together, could result in many new exciting qualitative findings, much as learning analytics has done in more traditionally quantitative fields. Furthermore—perhaps paradoxically—there are reasons to believe that qualitative methods and learning analytics can work particularly well together. A significant chunk of qualitative analysis has focused on listening to the data – building theory and hypotheses from data rather than prefiguring hypotheses before collecting those data. A similarly large portion of learning analytics concerns mining for patterns in data that are not immediately human-perceptible. By enabling these strands of learning analytics and qualitative learning sciences to mesh, and building tools that afford and support that mesh, we open up the potential for radical new understandings for and between the fields.

Finally, we suggested that this new meshed practice is more likely to be successful if it is supported by tools that are specifically tuned for this style of work. In that spirit, we presented the Tactic text mining environment as a tool that was designed with these aims in mind.

References
Empirical Evidence for Evaluation Anxiety and Expectancy-Value Theory for Help Sources

Iris Howley, Williams College, iris@cs.williams.edu
Carolyn Penstein Rosé, Carnegie Mellon University, cprose@cs.cmu.edu

Abstract: Expectancy-Value Theory for Help Sources (EVT-HS) states that whether or not students seek help from a particular source is determined by their perceived expectation that there will be help available, and the perceived value for the help from that source. This paper provides initial empirical validation for EVT-HS, while also introducing and providing support for costs of seeking help from a particular help source impacting intention to seek help. Our survey experiment shows that raising perceived expectancies and values for a help source significantly predicts a student’s intention to seek help. Our results also show that evaluation anxiety, as a potential negative value, inversely predicts intentions to seek help from a particular source.

Introduction
Appropriate help seeking is a necessary skill in becoming a successful self-regulated learner and it is highly correlated with student achievement in the classroom (Magnusson & Perry, 1992). Students who do not seek help with difficult concepts, or who fail to consult with instructors, or who request inappropriate help are not as likely to experience success as students who seek help effectively (Magnusson & Perry, 1992). Seeking help when necessary assists students in understanding complex concepts that they do not understand (Magnusson & Perry, 1992). However, the process of identifying a help need to actually pursuing that help is a complex path. Not all students successfully find their way.

In this article, we provide initial empirical evidence in support of Expectancy-Value Theory for Help Sources (Makara & Karabenick, 2013) as a lens for understanding what perceptions impact from whom students seek help. This theory encompasses many social and non-social factors that influence student decisions to seek help. With an understanding of the theory of help-seeking from help sources, designers of learning experiences can better support student help seeking. We designed a survey experiment to link perceived expectancies, values, and costs with intention to seek help from potential peer helpers. While the experimental context is intended to inform the future design of an online peer help support system, the experimental methodology is designed to allow for a broader interpretation of results.

Prior work
In this article, we focus on three steps of the help seeking process to investigate how learning science theory can be leveraged to improve learning environments. These three steps include: “decide to seek help from an external source”, “identify potential helpers”, and “implement strategies for engaging the helper” (Nelson-Le Gall, 1981). We focus on these three steps as the potential for immediate impact to reduce obstacles to effective participation is tremendous. These three steps also lend themselves well to a lens of Expectancy Value Theory for understanding how student beliefs impact their help seeking while we use a different theory, Expectancy-Value Theory for Help Sources (EVT-HS), to explain whether learners pursue potential helpers. EVT-HS is a more constrained interpretation of Expectancy Value Theory for learning, but instead of focusing on larger self-concept tasks, such as learning a concept, its focus is on the decision to seek help from a particular help source.

Help seeking in learning contexts
Nelson-Le Gall (1981) proposes one model of help seeking in which the student must:

1. first become aware of a help need,
2. decide to seek help from an external source,
3. identify potential helpers,
4. implement strategies for engaging the helper, and
5. reflect upon the help seeking attempt.

The first step toward help seeking revealed by Nelson-Le Gall’s task analysis is identifying a help need. If students have the metacognitive capabilities to monitor their progress and can detect when they encounter a problem, then it is possible for them to proceed to the next step in the help seeking model. However, if students are not aware that they have encountered an obstacle, then they will not seek help when necessary (Nelson-Le
Gall, 1981). Research shows that this metacognitive ability to identify a help need is developed through maturation and experience (Markman, 1977).

In order to make the decision to seek help, a person must first weigh the costs and benefits of doing so. Asking for assistance can help a student complete a task, but it can come with social and personal costs such as feeling less competent or receiving less credit (Nelson-Le Gall, 1981). Not everyone is equally as sensitive or aware of these costs and benefits. A student’s disposition and goals may also affect the choice.

Once a decision has been made to seek help, one must next select a helper. In this step, the decision is influenced by the student’s perceptions and knowledge of potential helpers as well as the social situation. These perceptions and situational factors include the sex and age of both the help-seeker and the helper, the role relationship and status of both parties, perceived willingness to help, perceived competence of the helper, and socioeconomic status (Nelson-Le Gall, 1981). These perceptions and situational factors are of particular interest because they can often be intentionally designed, especially within interactive learning environments in which the system connects students to the help they require.

Once the learner has decided to seek help, and decided on a helper, there are a variety of outcomes to expect, dependent upon the student’s goals in seeking help. Help-avoidance, executive (or expedient) help seeking, or instrumental help seeking (Nelson-Le Gall, 1981) are also similar to avoidant, autonomous, and dependent help seeking behaviors (Nadler, 1997). One can either ask for help or not, but one can also ask for help simply to complete a task quicker or to learn more. There is also a distinction to be made among help seeking, information-seeking, feedback-seeking, answer requests, or error checks (Puustinen et al., 2011). Beyond general categories, one can also examine help seeking based upon linguistic features such as the directness and politeness of the help being sought (Puustinen et al., 2011).

If the desired help is not acquired, students are then forced to reevaluate their strategies for obtaining help and may repeat the previous steps until help is achieved (Nelson-Le Gall, 1981).

Expectancy Value Theory
Students’ decisions to pursue learning goals are determined by their expectancies for success, and the values they place on the outcomes that come from that success. Eccles & Wigfield (2002)’s Expectancy Value Theory provides a larger model that includes these expectancies and values, but also incorporates students’ beliefs and self-schema. This model can be applied to a wide range of learning-oriented behaviors, including help seeking. We will be focusing on the direct antecedents that determine students’ achievement-related choices and performance: the expectation of success and subjective task value. We apply Expectancy Value Theory at the help seeking process level, and later introduce Expectancy Value Theory for Help Sources for understanding student behavior when selecting a help source.

Expectancy Value Theory for Help Sources is derived from Expectancy Value Theory, although due to its more targeted focus there are some differences. Both beliefs about the help source and more general beliefs about help seeking expectancies and values are important factors in the process of deciding to seek help. In this section we describe and distinguish Expectancy Value Theory from EVT-HS.

Modern Expectancy Value Theory
The Expectancy Value Theory (EVT) of modern educational psychology incorporates students’ ability beliefs, expectancies for success on a particular task, and four different task values (i.e., intrinsic, utility, attainment, and cost) (Eccles et al., 1983). More recent work on Expectancy-Value Theory has pointed toward expectancy for success being more predictive of performance and value beliefs being more predictive of achievement-related choice and effort (Trautwein et al., 2012).

Trautwein et al. (2012) provides evidence for the relationship between expectancy and value for a task to be enhancing. That is, in their model, both expectancy and value positively predicted performance and their interaction produced a stronger than additive effect on performance. These results held true in both mathematics and English language learning domains, which suggests that the interaction of expectancy for success and task values should be included as a term in Expectancy Value Theory analyses and models. Trautwein et al. (2012) also included costs of performing a task alongside intrinsic, utility, and attainment values which is similar to other work incorporating costs into Expectancy Value Theory (Eccles & Wigfield, 2002). While “low cost” was significantly correlated with the other values, the correlation between utility values and low cost were generally stronger. The authors hypothesized that low cost and utility values represent extrinsic values, while the attainment and intrinsic values were considered more intrinsic. However, this explanation assumed a particular definition of costs, largely focused on the amount of time required of the student to pursue that avenue of help.
Expectancy of success also has multiple dimensions in modern Expectancy Value Theory. These dimensions consist of broad ability beliefs about competence in a particular domain, and a more narrow expectancy of success on a specific upcoming task. Research has shown that these two dimensions are highly correlated, and in many real-world achievement situations they are empirically indistinguishable (Eccles & Wigfield, 2002). Trautwein et al. (2012) measured expectancy for success in math and English with a self-concept instrument with items such as “I am good at mathematics/English.” It is a common practice of modern research on Expectancy Value Theory to use similar self-concept items rather than expectancy for success items for a specific upcoming task. While Expectancy Value Theory has directed expectancies for success to be measured more generally at the self-concept level, it is important to include expectancies for success on a specific upcoming task into initial investigations of Expectancy Value Theory for help seeking.

**Expectancy Value Theory of help sources**

The process of pursuing help, from Nelson-Le Gall’s (1981) model, once the student has decided to seek help requires (3) selecting a source from which to seek help and then (4) to follow through with the help request. While the actual pursuing of the help can be understood through the Eccles & Wigfield Expectancy Value Theory model as previously described, the selection of a help source can be better examined through Makara & Karabenick’s (2013) Expectancy Value model of help seeking for help sources, below:

![Figure 1. Makara & Karabenick Expectancy Value Theory (2013) for help sources model.](image)

Expectations for help from a particular help source are based on beliefs about whether that source will be available to provide help, whether that source is accessible, and a basic belief that there will be obtainable help from that particular source. Values for a help source originate from whether that help source will be able to provide the expected type of help such as the expected quality and accuracy. This model functions as an initial theoretical explanation for how students select and seek help from a particular resource, but empirical support for this framing of expectancies for success and values is not yet empirically validated.

Makara & Karabenick (2013) define expectations for success from the help source as the belief there will be help which is more about the help source rather than student self-concept. A student might have general self-concept beliefs about their past successes in help seeking in general, but also self-concept beliefs for the pursuit of help from a particular help source. It is currently unknown how these two levels of expectations for success in help seeking combine to influence help seeking from a particular source.

Expectancy Value Theory for Help Sources also largely focuses on utility value for the values for the help source. This is likely due to the other, more intrinsically-related value types being less relevant in a help seeking concept. Intrinsic value was measured by items in Trautwein et al. (2012) such as, “I enjoy puzzling over mathematics/English problems.” It seems unlikely that students would enjoy seeking help and consider attaining help as a personal value. However, Makara & Karabenick (2013) could certainly have included cost items in their model of Expectancy Value Theory for Help Sources. While cost can be measured as an amount of time required to achieve the task as in Trautwein et al. (2012), it can also be measured as public and private threats to self-esteem, and inconveniencing a particular help source. Students can perceive costs at the help seeking level (i.e., “Others would think I was dumb if I ask for help in this class”, Wolters et al., 2003), but also at the help source level (i.e., “This helper would think I was dumb if I asked them for help in this class.”). While costs were not explicitly included in Makara & Karabenick (2013), an examination of costs’ function in seeking help from a particular source is an additional purpose of this article.

In determining from whom/where to seek help and whether to actually pursue that help, we must consider student expectations and values for the help from a help source. This examination should incorporate costs for seeking help as part of gaining a wider understanding of beliefs for a help source. Better understanding the relationship between expectancies and values for a help source and student self-concept beliefs about their help seeking is an additional goal.

**Costs of seeking help**

This section introduces costs of seeking help with a particular focus on social costs of seeking help. Costs are typically considered one of many possible values (alongside attainment, intrinsic, and utility values). But there
are also many different types of costs. Costly outcomes can include private threats to self-esteem (i.e., “If I ask for help, it means I’m not competent”), public threats to self-esteem (i.e., “If I ask for help the teacher will think I’m not competent”), face threatening acts (i.e., “It will inconvenience the teacher to help me”), among others. Costs were not included in the Expectancy Value for Help Sources Theory explicitly, but certainly one help source could induce more costs than another.

Evaluation anxiety, often referred to as evaluation apprehension, or a person’s concern about being evaluated (Guerin, 1986), can be impacted by numerous contextual factors and is also similar to perceived public threats to self-esteem (Shapiro, 1983). Both of these factors are related to impression management strategies to prevent others from perceiving one as incompetent. In this section, we focus on the effect of evaluation anxiety in learning contexts.

Anxiety related to the potential to be evaluated, whether implied or explicitly stated, is known as evaluation anxiety (Cottrell et al., 1968). Learning often requires evaluation, either from others such as the teacher or from within when self-monitoring one’s progress, and so the issue of anxiety around evaluation potential is relevant to learners. However, a review of the literature does not appear to reveal evaluation anxiety systematically studied with regards to its effects on help seeking. Evaluation anxiety is referred to, specifically in reference to its relationship with threats to public self-esteem, as in Nadler (1997) in which the author posits that threat to public self-esteem is an explanatory concept for participants avoiding seeking help on ego-central tasks, and that this “suggests that one avoids the seeking of help because of evaluation anxiety concerns.”

Modern methods to measure evaluation anxiety look not only at experienced evaluation anxiety, but also the cause of the evaluation anxiety (Leary et al., 1986; Bagley, 2007). To measure experienced evaluation anxiety, a subset of items used to measure negative affect are used as a scale. These items include negative affects specifically related to anxiety: nervous, worried, calm, tense, and relaxed (Leary et al., 1986).

**Empirical evidence for Expectancy Value Theory for help sources**

As an initial step toward understanding how evaluation anxiety and EVT-HS impacts student help seeking, we devised an online experiment testing the EVT-HS with evaluation anxiety model. This initial survey connects potential help source manipulations to items measuring expectancy value beliefs toward the help source, evaluation anxiety, and intention to seek help. The purpose of this experiment is to serve as initial empirical evidence for the relationships between expectancy for the help source, values for the help source, and help seeking outcomes as proposed in Makara & Karabenick (2013). This experiment also provides initial empirical evidence for including evaluation anxiety as a social concern in help seeking from a particular source.

Expectancies, values, and costs for the help source can be manipulated through the presentation of potential help sources. We are grounding the context of the help sources in a peer helper support system that our research group is in the process of designing and testing. Potential helper screenshots should directly manipulate perceptions of expectancy and value for help sources. The helper screenshot, as shown in Figure 2, is on a blue background with an anonymized profile image and username, and one of four possible sentences that represent our experimental manipulation:

1. “This person is a fellow student” (control)
2. “This person is available to give help” (expectancy)
3. “This person offers high quality help” (value)
4. “This person will evaluate the quality of your question” (cost)

By deriving these sentences directly from Makara & Karabenick’s (2013) EVT-HS, we intend to test whether the theory has the hypothesized effect on help seeking attitudes towards the help sources.

![Figure 2](image.png)

**Figure 2.** A Helper screenshot with the ‘values for help source’ manipulation sentence. Other sentences were used for the other three conditions.
Research hypotheses

Our research hypotheses are derived from the direct relationship between the EVT-HS theory, and our manipulations. There is a set of hypotheses dedicated to the relationship between the manipulations and the beliefs, and an additional set of hypotheses related to the beliefs and the help seeking outcome. Figure 3 presents our tested hypotheses and the associated significant relationships.

![EVT Helper Survey Experiment](image)

**Figure 3.** The hypotheses results model for the EVT Helper Survey Experiment. Black solid lines indicate supported hypotheses, grey solid lines are unsupported hypotheses, and black dotted lines are un-hypothesized relationships. Arrow-less lines indicate correlations.

Connecting manipulations to beliefs

1. The Expectancy Sentence (“This person is available to give help”) will increase self-reported expectations for the help source, more than the Control Sentence (and other sentence conditions).
   - *(Partial Support)* The Value Sentence and Expectancy Sentence resulted in significantly more self-reported Expectancy Beliefs for the help source, than the Cost and Control Sentences, F(3,159)=13.68, p <0.001, R² = 0.68.

2. The Value Sentence (“This person offers high quality help”) will increase self-reported values of the help source, more than the Control Sentence (and other sentence conditions).
   - *(Supported)* The Value Sentence predicts significantly more self-reported Value Beliefs than the Expectancy and Control Sentences which predict more than the Cost Sentence, F(3,159)=35.35, p<0.0001, R²=0.64.

3. The Cost Sentence (“This person will evaluate the quality of your question”) will increase self-reported costs for the help source, more than the Control Sentence (and other sentence conditions).
   - *(Supported)* The Cost Sentence significantly predicts more Cost Beliefs (i.e., evaluation anxiety) than the Expectancy and Value Sentences, with the Control Sentence being statistically indistinguishable from the Cost and Expectancy conditions, F(3,159)=2.80, p = 0.04, R²=0.75.

Connecting beliefs to intention to seek help

4. The EVT-HS beliefs should connect to help seeking outcomes.
   a) Expectancy for Help Sources Beliefs should significantly positively predict intentions to seek help, *Supported:* F(1,165)=401.11, p<0.0001, R²=0.79.
   b) Value for Help Sources Beliefs should significantly positively predict intentions to seek help, *Supported:* F(1,213)=245.77, p<0.0001, R²=0.77.
   c) Cost beliefs (i.e., evaluation anxiety) should significantly negatively predict intentions to seek help, *Supported:* F(1,212)=25.83, p<0.0001, R²=0.69.
   d) Expectancies and Values for the Help Source should interact as an enhancing model on the prediction of intention to seek help, *Partially Supported:* β = -0.03, t(200) = -1.69, p = 0.09, R²=0.86.

Study design and methodology

54 participants were recruited from a private American university’s participant pool. 7% of respondents experienced no college education, 20% received some college education, 43% obtained a bachelor’s degree, and 30% had a graduate degree. 61% of respondents were from the United States, 11% from East Asia, 24% from...
India, and 4% from Eastern Europe. Respondents had a mean age of $\mu=28.7$ years, $\sigma = 9.8$ and 63% of survey respondents were female. Each participant visited the online survey, read the instructions, viewed a helper screenshot, and then completed survey items measuring our constructs of interest. Each participant saw all four help source sentences randomized in this within-subjects survey experiment.

**Survey items**

Dependent measures were evaluation anxiety items from Leary et al. (1986), intention to seek and avoid help from the self-regulated learning literature in Wolters et al. (2005), and newly designed items we derived from the Expectancy Value Theory for Help Sources, shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Survey items derived from EVT-HS to measure perceived expectancies and values for a help source.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expectancy Beliefs for the Help Source</strong></td>
</tr>
<tr>
<td>1) &lt;This person&gt; is available to give me help.</td>
</tr>
<tr>
<td>2) If I ask for help from &lt;this person&gt;, they will give me help.</td>
</tr>
<tr>
<td>3) If I have a question for &lt;this person&gt; they will answer me.</td>
</tr>
<tr>
<td><strong>Value Beliefs for the Help Source</strong></td>
</tr>
<tr>
<td>4) The help from &lt;this person&gt; will be what I need to answer my question.</td>
</tr>
<tr>
<td>5) &lt;This person&gt; will provide me answers of high quality.</td>
</tr>
<tr>
<td>6) &lt;This person&gt; can give me accurate help.</td>
</tr>
</tbody>
</table>

Cronbach’s $\alpha$ for our developed measures of Expectancy Beliefs for Help Sources items was 0.93 and for Value Beliefs for Help Sources, $\alpha = 0.96$, indicating excellent internal consistency. Survey items for each experimental instance were combined into a single scale item representing evaluation anxiety, expectancies, values, and intention to seek help from the particular help source. Figure 4 shows the median and quartile values of the response scales as a boxplot.

The above survey items were preceded by a Helper Screenshot, as in Figure 2, which were in turn preceded by the following hypothetical framing: “You are enrolled in an online course and are having difficulty with one of the assignments. You decide to seek assistance from some of your peers. You submit your question to the online course website and the course system recommends you ask the following fellow student:”

**Statistical approach**

All analyses connecting categorical experimental manipulations to numerical beliefs scales were performed as an ANOVA with RespondentID as a random effect to account for the within-subjects experimental design. Since each participant saw all four of the sentence manipulations (control, expectancy, value, and cost), the categorical condition variable had three degrees of freedom. Analyses connecting the theory beliefs scales to intention to seek help were performed as a linear regression with RespondentID as a random effect as well. Post-hoc analyses connecting levels of variables to outcomes were performed via Student’s t-tests.

**Figure 4.** On the left, a box plot showing aggregate median responses and quartiles for survey scales. **Figure 5.** On the right, a line of fit with confidence of fit shading showing a significant interaction between values for the help source and evaluation anxiety on intention to seek help.
Results
In general, as shown in Figure 3, our hypotheses were mostly supported, except the Value Sentence manipulation impacted expectancy for the help source beliefs, just as much as the Expectancy Sentence manipulation did. Statistical relationships are reported underneath the hypotheses in the Research Hypotheses.

There was also a significant interaction between the evaluation anxiety and [positive] values for the help source variables on intention to seek help, $\beta = .12, t(210) = 2.47, p = 0.01$, as shown in Figure 5. As the values for the help source rise, the perceived evaluation anxiety caused by that help source decreases, although less steeply. This suggests that when students believe a helper will provide good quality help, they are also slightly less afraid of being evaluated by that helper, therefore evaluation anxiety for the help source may very well be functioning as a negative value belief.

Trautwein et al. (2012) suggests that expectancies and values in a domain-centric Expectancy Value Theory model should contain an interaction between expectancy and value with an enhancing effect. A marginal interaction between expectancies and values for the help source was found, $\beta = -0.03, t(200) = -1.69, p = 0.09$, $R^2=0.86$. However, if the interaction term had an enhancing effect, the lines in Figure 6 would fan outwards. Instead, only some of the lines fan. It appears that the value beliefs for the help source hit a 1.73 standard deviation maximum, and a ceiling effect occurred, preventing a clear enhancing interaction from occurring.

![Figure 6. Expectancy and Value predicting Help Seeking shown as standard deviations from the mean.](image)

Limitations
This survey experiment employed the use of a hypothetical online classroom and assumed the participant had a hypothetical question to ask in order to provide context for selecting a helper. While our results provide some confidence in the environmental validity of this method, the ecological validity might be questionable. Furthermore, each manipulation only had one sentence, and so respondent beliefs might be in response to the phrasing of the question and not the larger theory manipulation the sentence was designed to represent. Response bias is also a possibility in survey experiments, and so future work includes replicating our results.

We were unable to manipulate Value Beliefs for the Help Source separately from Expectancy Beliefs for the Help Source. A few possible explanations consist of: (1) the Expectancy Beliefs items could be inaccurately constructed, (2) the value manipulation sentence was inaccurately constructed, or (3) the Expectancy Value Theory for Help Sources requires refinement. It may not be possible to realistically manipulate value beliefs separately from expectancy beliefs for the help source. Future work should investigate the design of manipulatives that can impact values for the help source beliefs separately from expectancies.

While an enhancing interaction between expectancies and values for the help source was expected, the marginal interaction did not fulfill this enhancing relationship. A ceiling effect is likely a partial explanation.

Future work
While this survey experiment provided strong evidence in support of EVT-HS and evaluation anxiety as impacting intention to seek help, next steps involve a second survey experiment using a different help seeking context to provide for more generalizability. Also, investigating these constructs in a live learning environment could reveal EVT-HS’s environmental validity. We have a peer helper suggestion system for online courses to which we plan to apply this newfound understanding of help seeking. Future work should also include the development of more sentence manipulatives to better operationalize the effect on EVT-HS beliefs.
Conclusion
From our results, we see that the Expectancy-Value Theory for Help Sources impacts self-reported help seeking outcomes. Our survey experiment shows that raising perceived expectancies and values for a help source significantly predicts a student’s intention to seek help. Our results also show that evaluation anxiety, as a potential negative value, inversely predicts intentions to seek help from a particular source. This initial evidence supports Expectancy Value Theory for Help Sources in its proposed form.

However, value beliefs for the help source are difficult to manipulate separately from expectancy beliefs. A student could believe that if resource A gives good quality help, then they probably are available to give help. From a practical standpoint, this is less of an issue, as course designers and instructors can focus on creating a learning environment that raise either (or both) expectancies or value beliefs for the help source to garner a positive impact on help seeking. From a theoretical standpoint, the fact that values cannot be easily manipulated separately from expectancy beliefs for the help sources suggests that Expectancy Value Theory for Help Sources might require some refinement to be more useful in explaining the specifics of seeking help from a particular source. The hypothesized relationship between our measures of expectancies, values, and costs to intention to seek help was supported, which implies that it is not purely a question of theory, but a matter of manipulating the theory beliefs in a real-world setting.

Overall, in this paper, we have provided initial evidence in support of expectancies, values, and evaluation anxiety beliefs as influencing whether people seek help from a particular helper. Our results also emphasize that learning experience designers incorporate awareness of student evaluation anxieties into their courses, as it can impact help seeking as well.

References
Abstract: Although efforts to critique deficit models of cognition and address educational inequity are foundational to the learning sciences, there has been particularly rich work done in this tradition in recent years. In this paper, I offer a set of pathways that characterize this work towards justice in the learning sciences, work which reflects a variety of means and ends. The pathways are: (1) Consider the goals of an equity-oriented framework for learning; (2) Theoretically, draw on existing critical social theory; (3) Methodologically, focus on collaborative change-making, and; (4) Support heterogeneity in knowing and doing. These pathways are neither exhaustive nor mutually exclusive, but rather provide an assessment of the current justice-oriented landscape in the learning sciences. I then argue that heterogeneity in means and ends towards addressing learning inequities is fruitful for the field and offer some open questions with which the critical learning sciences has yet to fully contend.

Keywords: equity, dignity, justice, heterogeneity

Introduction

At the 2014 ICLS Conference, Booker, Vossoughi, and Hooper (2014) advocated for a political theory of learning. They compellingly argued, “If we are interested in what the learning sciences can offer for how to democratize learning, we need our theoretical and methodological tools to help us wrestle with inequity and dehumanization” (p. 921). Although these ideas are not brand new to the learning sciences (LS), we as a field are at but the onset of our theorizing about the relationship between social (in)justice and learning (Politics of Learning Writing Collective, 2017; Nasir, Rosebery, Warren, & Lee, 2014). In the four years since Booker et al.’s call, a significant amount of LS work towards educational equity has emerged. This includes a special issue of the Journal of the Learning Sciences on social justice (Tabak & Radinsky, 2014), a special issue of Cognition and Instruction on participatory design research (Bang & Vossoughi, 2016), and a new volume on the intersection of critical social theory and sociocultural learning theories, Power and Privilege in the Learning Sciences (Esmonde & Booker, 2017). This work together recognizes the potential LS has to combat discriminatory social structures, empower all learners, and bring positive social changes.

In this paper, I review many of these recent contributions with the goal of understanding the nature of contemporary learning scholarship that focuses explicitly on issues of equity, justice, and learner dignity. As I will discuss, what equity-oriented research means has varied considerably, but in general this work shares a focus on broadening participation in learning and eliminating discrimination in educational environments. I explicate four pathways that characterize how LS scholars have thought through equity philosophically, theoretically, and methodologically. I offer the pathways as a heuristic for thinking through the flavor of LS research on equity to date. It is my hope that the pathways outlined here: (a) clarify the shape and form of equity-oriented research that informs the field broadly, and (b) provide a starting point for researchers new to such approaches to find resonances between their research and existing work.

Given space limitations, this paper is not meant to be a comprehensive review of the literature. The initial readings discussed above led me to new authors and new readings. As I engaged in identifying and reviewing the relevant literature, I employed a broad understanding of “learning sciences research” to include the general study of how people learn. In addition, I drew upon other literature when it seemed to inform the pathway being discussed. Of course, this resulted in the inclusion of some literature at the expense of other pieces. However, I do not mean for these pathways themselves to be exhaustive or mutually exclusive; rather they are incomplete and intersecting. The pathways are meant to call attention to the multiplicity of ways that learning sciences researchers have and might orient to scholarship in the service of equity. I conclude by discussing the importance of heterogeneity in our approaches to criticality and offer some open questions for the field to consider moving forward.

Pathway #1: Consider the goals of an equity-oriented framework for learning

Philip and Azevedo (2017) outlined four discourses in science education research that seeks equity. The first two discourses focus on helping students connect to and identify with science, and in so doing seek to make sure learners have “a seat at the table.” The third discourse seeks to leverage out-of-school science to reconceptualize what is valued in in-school science, and so it focuses on changing the discipline to be less hostile to learners. The
fourth discourse focuses on how science is and is not used to achieve justice through activism on the larger social level, and so this research focuses on participation in work to dismantle oppressive social systems. The authors stress that the third and fourth discourses, which center the transformation of systems rather than focusing on individuals, are not necessarily better than the first two, which may make available more immediate material gains for learners. Likewise, in LS writ large, we need not make a hierarchy of equity-oriented research. While “big picture” research that thinks deeply about the transformation of social systems is necessary, so too is research that broadens participation even in unjust systems (e.g., helping minoritized learners access profitable career paths) that results in material gains for minoritized individuals. It is promising that heretofore LS research has supported a variety of definitions of what “equity” looks like and means.

Any understanding of equity and learning must resist the belief that more learning is necessarily a means towards equity. Learning is not a priori a good thing; many people learn things like racist practice or ineffective folk theories of teaching. This knowledge is dangerous insofar as one person’s learning of such knowledge has negative symbolic and material consequences for other people. This is summarized well by Biesta (2013):

To suggest that learning is good or desirable – and thus to suggest that it is something which should go on throughout one’s life or which should be promoted in schools – does therefore not really mean anything until it is specified what the content of the learning is and, more importantly, until it is specified what the purpose of the learning is. (p. 6).

Indeed, as Biesta notes, when we focus on learning-as-change, we – perhaps unconsciously – make an evaluative judgment that the temporally earlier state of affairs was bad or worse than the amorphous “after learning.” It is possible that this focus on learning-as-change might cause us to approach an individual as damaged and the learning process as one of repair (a deficit orientation). That said, oppressive aspects of social systems must be challenged and subject to change. Therefore, equity-oriented work on learning might avoid deficit perceptions of individuals by choosing to focus on structural, social, and community entities.

Equity-oriented work that centers individuals might also ask how the learner and others are benefitting from the individual’s engagement in learning practices. Undoubtedly, certain kinds of learning better position learners to perform well on standardized tests, be admitted to prestigious higher education institutions, and access particular career opportunities. Scholars may additionally believe there is something intrinsically powerful about the disciplinary skills and content we seek to help learners acquire. Contention with these ends of educational equity represents the first pathway of equity-oriented research learning. It asks: why learning?

Espinoza and Vossoughi (2014) conceptualize the learning process as one in which learners can realize their human dignity and potential. Grounding their analysis in a history of learning among African American people in the United States – who through slavery and its compatriots were violently and explicitly forbidden to learn – they write:

Insofar as learning helps persons and selves flourish, it is dignity-conferring. Dignity can be derived from productive participation in the process of learning. Although perilous, it can also be acquired from resistance to the inaccessibility of the opportunity to learn. (p. 287).

A focus on human dignity challenges us to consider the relationship of learning to legally-sanctioned violations of human dignity (e.g., modern mass incarceration in the United States). With this in mind, scholars have asked what learning provides to the learner that does not come at the cost of dehumanizing them (e.g., by “pushing” learners through inequitable environments). For example, a career-oriented focus on STEM learning pushes women through the “STEM pipeline,” at the end of which they still make significantly less money than their white male counterparts (Sengupta-Irving, 2015). Instead, Sengupta-Irving (2015) argues that math education can (but does not automatically) offer an opportunity for self-determination and for becoming. Vossoughi (2015) proposes the notion of intellectual respect as key to equitable environments. In practice, an example of such an approach is found in the work of Gutiérrez (2008), in which youth are empowered to learn sociocritical literacy, tools to support their individual and collective moves towards justice. The program is explicitly designed to support deep academic learning – that directly opens college and career pathways – while ensuring learners can engage in social dreaming, seeing themselves as cultural-historical actors with agency to create a better world. Bang (2017) points out that a decolonial and indigenous ethic towards sociocultural learning theory may offer great possibility for human and community collective continuance, that is, “a community’s capacity to be adaptive in ways sufficient for the livelihoods of its members to flourish into the future” (Whyte, 2013, p. 518). Loaded with possibility, these approaches conceptualize learning as an endeavor that brings learners into contact with powerful tools and discourses that unseat their individual and collective social oppressions. This pathway thus disturbs a conception
of the learner as human capital in a fundamentally capitalist, unjust economic system (Sengupta-Irving, 2015), and troubles a conception of the learning sciences as moving people through such systems more efficiently. As Philip, Bang, and Jackson (2017) point out, thinking through equity involves asking not just how people learn but also “for what,” “for whom,” and “with whom” learning takes place.

Pathway #1 supports a deep fundamental reflection on what equitable learning moments, contexts, and systems look and feel like. This reflection is philosophical, cultural, social, and historical, but it has immediate implications for learning theory, design, and method, as discussed next. Scholars traversing this pathway help us tease out our evolving goals in justice-oriented learning sciences research. They also invite us to critique shallow conceptions of equity that stop at minoritized learners getting access to the otherwise-unaltered factory-model classroom. Instead, we too can engage in social dreaming about what an equitable world looks like and how supporting learning can help us get there.

Pathway #2: Theoretically, draw on existing critical social theory

Inequitable educational practices take place and are reified at multiple levels: some are societal, national, and cultural while others are quite local, happening in the classroom or out-of-school learning environment. The dimensions along which inequitable practices are organized of course varies from context to context. In the United States, for example, inequities persist towards individuals based on their race, religion, sexual orientation, gender, age, immigration status, dis/ability status, and other characteristics that should not be grounds for closing off learning possibilities. Critical social theories include general theorizations of power and oppression as well as feminist theories, critical theories of race, queer theory, critical dis/ability studies, decolonial studies, and others. Drawing on extant critical social theory is fruitful for at least three reasons. Firstly, understanding the complex cultural-historical roots of particular oppressed and power positions is necessary to create equitable learning environments that recognize the shared histories of particular groups without essentializing or stereotyping learners. Secondly, while pluralism, multiculturalism, and diversity are key values, they are abstract; drawing on critical social theory helps scholars explicitly name the kinds of injustice they are targeting (e.g., racism). Finally, employing specific social theory can be a way to address inequities across cultures, as discrimination does not take equivalent forms with equal salience across contexts.

LS has long theorized learning as a contextually-influenced activity. From a hierarchical understanding of context, critical social theory reveals that learners are not positioned by macro-contexts to have equal access to or be equally successful in learning micro-contexts. Moreover, a dialectical notion of context (e.g., Cole, 1996) reveals that inequitable learning moments are not merely consequences of societal injustice; such moments are also constitutive of societal injustice.

Critical social theories are seeing increased use in our field. McWilliams (2015) explicitly drew on queer theory to understand how fourth- and fifth-grade students engaged with understandings of gender beyond a simplistic binary. Curnow (2016) used feminist standpoint theory to make sense of how individuals of color participated in an activist community and to critique the notion of “legitimate peripheral” participation when factors such as gender and race create discriminatory membership boundaries. Philip, Gupta, Elby, and Turpen (2017) draw upon Butler's (2009) notion of the ungrievability of lives and Ahmed’s (2004) notion of “stickiness” to understand the participation patterns in a higher education classroom discussion on the ethics of drone warfare. Critical Race Theory (CRT; e.g., Ladson-Billings, 1998) has recently seen productive use in teacher education scholarship in the learning and cognitive sciences (Larkin, Maloney, & Perry-Ryder, 2016), although its potential has yet to be fully leveraged in LS (Parsons, 2017). Pathways forward also exist that draw on poststructural race theory (e.g., Shah & Leonardo, 2017), critical dis/ability studies (e.g., Smagorinsky, Cole, & Braga, 2017), and critical theories of race (Nasir & Hand, 2006).

In general, this pathway is characterized by explicit use of critical social theory to understand power and oppression at local and systemic levels (Esmonde & Booker, 2017). This scholarship appears particularly productive when it seeks to understand power generally as well as specifically with regard to particular demographic identities. Pathway #2 clarifies our understanding of how people learn while carefully avoiding cultural stereotypes, opting instead for a robust treatment of identity that considers historical, legal, and contemporary dehumanizations and discriminations towards particular groups. This pathway also offers potential to “speak back” to the development of these critical social theories by advancing their rich theorizations of learning and knowing processes.

Pathway #3: Methodologically, focus on collaborative change-making

Given the field’s methodological diversity and the damage-centered harmful methods that have characterized much social science research (Tuck, 2009), methodology is a productive area around which to organize equitable scholarship. While by no means the only organizing framework for LS research, a hallmark of the field is design-
based research (DBR; Design Based Research Collective, 2003). DBR draws on design experiments (Brown, 1992) to understand the efficacy – and potential efficacy – of well-constructed learning activities and environments. Extending and modifying DBR has been a fruitful starting point for scholars seeking to develop robust methodologies that are explicitly focused on understanding how people learn while also centering sustainable change-making and seeking equity. LS methodologies in service of justice need to fulfill at least two criteria: they must (a) inform our study of how participants learn about or through equitable practice and (b) must themselves be equitable research practices. Although we can never disregard the importance of the researcher in enacting research that is just towards and directly benefits participants, these approaches may be particularly effective at fostering critical reflection on the part of the researchers and provide some ethical guidance.

A focus on sustainable, systemic change-making characterizes methodologies that involve relevant stakeholders who have particular authority to influence learning environments. Design-based implementation research (DBIR) is one way to do this, as it focuses on bringing together insights from implementation research and organizational change with those of the learning sciences (Penuel, Fishman, Cheng, & Sabelli, 2011). Related is the conceptualization of research-practice partnerships (RPPs; Coburn, Penuel, & Geil, 2013), which are “long-term, mutualistic collaborations between practitioners and researchers that are intentionally organized to investigate problems of practice and solutions for improving district outcomes” (p. 2). DBIR and RPPs have seen notable success by working towards change through sustainability in formal schooling settings; as approaches, they bring together those who are accorded power in educational settings (e.g., teachers, superintendents) with researchers to support these changes in organizational cultures.

In some contrast, another common approach to change-making in education research focuses specifically on youth as stakeholders. Participatory action research (PAR; e.g., Cammarota & Fine, 2010) is an approach to research that foregrounds participants’ involvement in the research process, including questions asked, data collected, and resultant action. For example, Kirshner, Pozzoboni, and Jones (2011) took up youth participatory action research to study how youth learn to manage bias (conceptualized as openness to disconfirming evidence) in a collaborative way. PAR can be a tool to ensure that research participants are represented in and benefit from the research. Clearly, participatory approaches are not a panacea – they do not by themselves equate to just practices (Huf & Dennis, 2016). However, when PAR approaches focus on youth as agentic and reimagine systems to empower them, PAR may be a compelling pathway to more equitable research practices.

These values have emerged in other related approaches to DBR. Some scholars have focused on social design experiments (e.g., Gutiérrez & Vossoughi, 2010), which employ DBR approaches while also explicitly seeking social transformation by “creating a significant reorganization of systems of activity in which participants becoming designers of their own futures is an essential aim” (Gutiérrez & Jurow, 2016, p. 566). Still others have focused on an adaption of “participatory design” as used in human-centered computing scholarship (DiSalvo, Yip, Bonsignore, & DiSalvo, 2017). Bang, Faber, Gurneau, Marin, and Soto (2016) conceptualized their work as community-based design research, “a reworking of design-based research methods because it privileges and centers the work in community, engages broad ranges of community members, and is driven by community members in key project staff positions” (p. 31). These approaches focus on research as collaborative and empowering.

A coalescing of these approaches is represented by the emergent conceptualization of such research as participatory design research (PDR). PDR draws on PAR, community-based design research, social design experiments, DBIR, and more. PDR ultimately “maintains a commitment to advancing fundamental insights about human learning and development through explicit attention to what forms of knowledge are generated, how, why, where and by whom” (Bang & Vossoughi, 2016, p. 175). PDR is an emergent “big tent” for methodologies that seek equity through sustainable social change and learning that is collaboratively constructed and so collaboratively meaningful. PDR may be one of many potential pathways to counteract the harmful, damage-centered research that characterizes much social sciences research and particularly social science research on marginalized populations (Tuck, 2009). How to conceptualize these commitments is still a work in progress for this scholarship. Indeed, “It might be more appropriate to characterize PDR in its current state of development as an emerging field of research rather than a methodology (not to speak of method)” (Gutiérrez, Engeström, & Sannino, 2016, p. 281).

The vastness of these approaches can be overwhelming. A focus on their differences, while important in conceptualization and practices, may obscure what they have in common: equity-oriented research cannot be done alone. Our research should be with rather than on folks (O’Kane, 2000), should take seriously the people it purports to benefit, and requires a significant emotional and temporal investment on the part of multiple parties. This pathway has potential for supporting equitable practices and simultaneously adding to our knowledge about how people learn through scholarship that is deliberate, reflective, and slow (Mountz et al., 2015).
Pathway #4: Support heterogeneity in ways of knowing and doing

LS has long supported a multiplicity of ways of knowing and doing. For example, constructionist theory, rooted in a constructivist critique of objectivity in science, focuses on facilitating learning environments that promote epistemological pluralism, “accepting the validity of multiple ways of knowing and thinking” (Turkle & Papert, 1990, p. 129). Recent work continues to value these ideas, suggesting that cultural and epistemological heterogeneity grows possibilities to learn and that such heterogeneity is a foundational consideration in learning theory and design (Rosebery, Ogonowski, DiSchino, & Warren, 2010).

This approach to criticality focuses on broadening participation by thinking about the ways, tools, and participation patterns of knowing in dominant discourse, and working to unseat that dominant discourse by creating counter-narratives. This line of work seems particularly promising in STEM education, as scholars work towards projects in which access to ideas is mediated by thoughtfully-designed cultural tools that critique existing biases. For example, Pinkard, Erte, Martin, and McKinney de Roysten (2017) created the Digital Youth Divas project, in which youth co-designed new narratives about science and technology that facilitated the engagement of girls from non-dominant backgrounds in digital making. Similarly, Beuchley, Peppler, Eisenberg, and Kafai (2013) illustrate that the use of electronic textiles introduced conductive thread and fabric to circuitry learning in a way that broadened access for women and improved conceptual understanding for all genders. These approaches work in productive conversation with conceptions of identity as a joint accomplishment between a learner and their interaction with situated factors such as cultural tools (Hand & Gresalfi, 2015) and might support more equitable disciplinary identification (e.g., Bell, Van Horne, & Cheng, 2017).

In a related line of work, Medin and Bang (2014) have pointed out that a diversity of perspectives, approaches, and scientists results in higher-quality science across the field. However, “Underrepresentation in science will never be remedied by better schools, better curricula, better teachers, and all other betters that leave science itself as pure and beyond examination” (p. 240). Indeed, these approaches may be most effective when they focus on heterogeneity in people, tools, and discourses that are productive for learning in addition to fundamentally reexamining and broadening our understandings of disciplinary fields. This means encouraging the peer-supported and interest-driven practices that youth engage in and making visible the connection between these spheres and academic knowledge, as the connected learning initiative seeks to do (Ito et al., 2013). These connections must be dialectical, involving shifts in how we think about individual interest and how we think about academic knowledge.

This pathway is particularly promising for the proximal goals of increasing learner participation in disciplines that have historically been hostile to them, which may in turn benefit the learner materially through more robust college and career prospects. A reconceptualization of tools and narratives is helpful because such an approach tends to avoid deficit perspectives by focusing on factors that push out or limit participation. These approaches are consonant with the larger projects of the learning sciences, which views learners’ prior social and cultural experiences as assets that themselves constitute learning practices and in turn can be leveraged to promote academic and institutional learning if this is the desired goal (Gutiérrez & Rogoff, 2003; Lee, 2001; Rose, 2004). Despite this, it is easy to stereotype large groups of people by assuming their presence in a space by itself creates heterogeneity without attending to the participation structures that also circulate in that space. Further, effective work on reconceptualizing dominant tools, discourses, and practices must avoid assuming the same ways of thinking, being, and doing are familiar or appealing to each member of a minoritized group, as this erases the vibrant heterogeneity within groups. Indeed, as Esmonde (2011) warns, even nuanced academic research risks this essentialism if it fails to understand at a deep level how these social categories of interest are constructed and performed. Focusing on specific critical social theory (Pathway #2) and designing with learners (Pathway #3) may help mitigate some of these risks.

Discussion: Heterogeneity and quality in the critical Learning Sciences

It is my view that the support of epistemological and practical heterogeneity (Pathway #4) must apply not only to the learners with whom we work, but to our own research and to the field of the learning sciences. Medin and Bang (2014) argue, “Assuming that scientist diversity is correlated with diversity in methods and theoretical orientations, we have a compelling reason to believe that scientist diversity makes for better science” (p. 234). I believe this claim extends to the learning sciences, which surely benefit from heterogeneity in theory, method, and participants. As this paper has tried to show, there is not one “right” way to do equity-oriented learning research. Rather, there exist multiple pathways that have proximal and distal implications for unseating dehumanizing power structures. I believe it is possible to traverse all four pathways simultaneously in a research project or program in its theory, method, and designs; it is an open question if the pathways can be meaningfully disentangled, and the advantages and disadvantages of traversing one without the others remain unknown.
This said, traversing these pathways is not a guarantee that research is equitable or in service of justice. There is always a risk of reifying the very structures we seek to challenge if we are not reflexive about who is benefiting from our research and what high-quality equity-oriented research looks and feels like. Calling for heterogeneity is “not advocating an anything goes approach. To the contrary, this perspective advocates for intentional engagement of the heterogeneity that is ubiquitous in classrooms” (Rosebery et al., 2010, p. 327) and in our research. Quality always matters, and an attention to equity is by no means the only way, let alone a guarantee of, research quality.

We have yet to fully conceptualize validity and quality across these threads. In my view, the general validity criteria from our specific methods can sometimes still be applied to equity-oriented learning research. In other places, openly ideological research may require ideas less frequently discussed in the field, such as what Lather (1986) and others refer to as catalytic validity, or “the degree to which the research process re-orient, focusses, and energizes participants in what Freire (1973) terms ‘conscientization,’ knowing reality in order to better transform it” (p. 67). Productive conversations across perspectives within equity-oriented learning research will surely be helpful for thinking about the quality of the work, just as it has been in educational research more generally (e.g., Moss et al., 2009). I conjecture that the language of “quality” will help us clarify what we as a field believe our trajectory should be. As Philip and Azevedo (2017) argue, “While all equity-oriented work is important, not all equity-oriented work is equal” (p. 525). Much more needs to be done to support how we think through effective critical learning sciences scholarship. What need it do for participants? What need it contribute to our knowledge of how people learn? How should it account for researcher positionality? A diversity of approaches and researchers will surely deepen this conversation.

Conclusion

Given the growth of critical learning sciences research in recent years, it is possible to read this paper as a critique of work in the field that does not fit into these pathways. To some degree this is reasonable, but I do not intend these pathways to be an exhaustive taxonomy of such research; indeed, much high-quality equity-oriented research does not (and should not) fit neatly into one of these dimensions. Furthermore, a focus on criticality and justice is not a critique of the learning sciences writ large. Indeed, scholars who do equity-oriented work have used core ideas from the learning sciences to develop their approaches. Thus, the learning sciences has a role to play in lending something quite fruitful to equity-oriented goals. With this in mind, a focus by many on equity should not be interpreted as an indictment but rather as a generative critique (J. Lester, personal communication, November 14, 2017) of the field.

As I have demonstrated, the study of injustice and learning is being done robustly and is even foundational to the learning sciences. But it is my hope that LS has only just begun to contend with the role of justice in the learning process. For example, more work should be done to test if our theoretical infrastructure is robust enough to account for the deep historical injustices that surely color learning contexts. For the moment “it is not clear what the affordances are in shifting from a sociocultural to a sociopolitical analysis of learning […] versus employing a more expansive sociocultural analysis that attends to the sociopolitical as well” (Gutiérrez, Engeström, & Sannino, 2016, p. 278). We might also continue our conversations about quality and heterogeneity in equity-oriented learning research. As scholars continue to take these pathways and as they increasingly intersect, we may indeed develop newer theorizations of not only when and where people learn, but the very mechanisms of how people learn. These pathways forward are promising, for, just as in the early days of the field, “The failures are sure to outnumber the successes by a goodly margin, making it certain that [we] will never run out of interesting things to do” (Cole, 1996, p. 350).

References


Sengupta-Irving, T. (2015, November). “Tell them we are rising: Serving more than a pipeline through humanistic STEM learning.” In S. Vossoughi, C. Malsbary & F. Erickson (Chairs), Revisiting the past to shape the future. Proceedings of the 114th American Anthropological Association Meeting: Denver, CO.


Expanding the Maker Movement by Recentering “Building for Others” in Construction Activities

Nathan Holbert, Teachers College, Columbia University, holbert@tc.columbia.edu
Sawaros Thanapornsangsuth, Teachers College, Columbia University, st2839@tc.columbia.edu

Abstract: The goal of the Bots for Tots project is to develop theory around the role of activity framing on diverse participation in making and construction practices. In Bots for Tots, a design-based research project in three locations over two years, builders create a “dream toy” for a younger member of their local community. We hypothesized that building for others would appeal to young girls that may not typically see building robots and racecars as aligning with their goals and values. Data indicates participating girls consistently considered their “client” throughout the design process and were interested in continuing to build for others after the project ended. We show how the degree to which the “other” impacts the construction process depends on the quality of the relationship between builder and client. We propose recentering “women’s ways of knowing” in maker practices to broaden participation among women in STEM domains.

Introduction
The “maker movement” has become an international phenomenon. No longer just the purview of a handful of technology enthusiasts, the opportunity to design and produce unique artifacts using high tech prototyping equipment such as 3D printers, laser cutters, and microcontrollers has extended beyond the garage and university laboratory. Schools and educational programs have embraced so-called makerspaces in hope of leveraging these tools and activities to engage young people in STEM domains and practices (Halverson & Sheridan, 2014). Despite this exciting development, longstanding traditions of “making”—such as sewing and textile work that has traditionally been performed by women—have mostly been ignored in the mainstream face of the movement (Buechley, 2016), leaving many to question whether the embrace of the maker movement by educators will help to address or exacerbate the overwhelming gender gap in STEM domains. Enthusiastic about the cognitive and sociocultural possibilities enabled by construction activities (Papert, 1980), we propose that when making is framed as leveraging multiple epistemologies, ways of knowing, and diverse materials and practices it can be a powerful method of connecting diverse learners to STEM domains.

Theoretical framework
Despite an explicit focus by education researchers, practitioners, and policy makers, women make up a small segment of the STEM workforce. Recent surveys suggest women hold just 25% of computing occupations and make up only 15% of the engineering workforce (NSF, 2015; NCWIT, 2015). Many factors contribute to this gender gap including a misogynistic culture that pervades STEM workplaces (Posner, 2017; Ratcliffe, 2015), the low number of female role models and support (Margolis & Fisher, 2003), as well as a misalignment between the values and goals held by many women and their perceptions of the work of STEM domains (Diekman et al., 2010; Intel, 2014).

As maker activities have become a popular method of engaging young learners in STEM practices, it is important to make clear that the popular “maker movement” has major diversity issues. For example, 70% of those attending the 2014 Maker Faire in San Francisco were men, an overwhelming 97% had college degrees, and attendees had a median household income of $130,000 (Maker Media, 2014). Examining the 53 covers of all Make: magazines published over 10 years, Buechley (2016) found that only 16% showcased women or girls.
One promising avenue for addressing the recent and narrow definition of making popularized by *Make:* has been to recenter making practices such as sewing and paper crafts that have existed for centuries and have traditionally been performed by women. This effort has included the design of new tools and technologies such as the Lilypad Arduino that brings computational power to textiles (Buechley, 2006), as well as a diverse array of workshops and activities that engage young women in designing electronically embedded clothing (Kafai, Fields, & Searle, 2014; Peppler & Glosson, 2012). Materials and tools are cultural artifacts that communicate gendered expectations. By expanding the possibilities of what these tools and materials can do, we expand what counts as legitimate STEM activity.

This work has pushed the research community to look beyond the overtly masculine definition of STEM to include the unique practices and values of the visual and manual Arts. And yet, too often the important intellectual contribution of these programs of research are oversimplified in practice and reduced to robots for boys and fashion and e-textiles for girls (Kafai et al., 2014; Kafai & Peppler, 2014). Consequently, our work has sought to further broaden this design space by considering not just the materials and tools of construction, but also the underlying values and motivation at the heart of making. Not just “who makes and with what,” but also “why and for whom?”

We find Belenky et al.’s (1986) work on women’s ways of knowing a compelling framework for reframing the value and goals of making to be about building connections. They argue that women are often driven by the desire to connect with knowledge and with “the other” at a personal level—that “connected knowers learn through empathy” (p. 115). Though these values are not restricted to women (and some women may value traditionally “masculine” activities and practices), these ways of knowing are ignored by STEM practitioners or worse, actively disparaged and belittled (Posner, 2017; Ratcliffe, 2015). Evidence of the value of reframing these activities to focus on contributing to one’s community can be found in the project-based service learning literature. Schools of engineering that have adopted project-based service learning have found greater success in both recruiting and retaining women and other underrepresented groups (Barrington & Duffy, 2007; Swan, Paterson, & Bielefeldt, 2009).

**Methods**

The BfT project was designed to engage young makers (builders) in building toys for other children in their community (clients). This project was conceived of as design-based research (The DBR Collective, 2003) and implemented iteratively over two years in three different locations, with each iteration evolving to address unique challenges and contexts and in response to ongoing data analysis. The design-based research methodology allows us to compare iterations of the BfT design across each version and between contexts thereby informing theory at the intersection of the design and student learning.

In the first year of the project, all 4th grade students (ages 9-10) from a public elementary school in a large urban city in the US were invited to attend a series of five free Saturday workshops where the goal would be to “build new toys for younger kids in your school.” These workshops were run by the first author and were held in a maker lab in the first author’s academic institution. While all 4th grade students received the flyer advertising the workshop, due to space constraints only the first ten to respond and schedule a pre-workshop interview were invited to participate. One participant that was interviewed was unable to attend the workshops resulting in nine participants (seven girls and two boys). In this paper, we will refer to this implementation as “Y1 Public.”

The following year, BfT was re-designed, in collaboration with a teacher partner, responding to findings from our first implementation (extensive details about the design changes can be found in Holbert et al., 2017). In one of these implementations, BfT was once again presented to 4th grade students from the same public elementary school as in Y1 Public. However, in this implementation, BfT was structured as an afterschool club that met weekly for 1.5 hours during a semester resulting in 12 total sessions. Because afterschool clubs are often used as childcare, and the school controlled the sign up process, it is likely that parents signed their children up to participate, rather than the students choosing to participate based on interest. This sign up process resulted in eight total participants (six boys and two girls). The BfT afterschool maker club was located in the school’s library and was facilitated by the second and first authors. We will refer to this implementation as “Y2 Public.”

In the second year, we also implemented BfT in an all-girls private school in a nearby suburban community. This implementation was run by our teacher co-designer who is employed by the school as the visual arts instructor and was held in the school’s expansive “Design and Engineering Lab.” All 4th grade students were required to participate in the BfT project resulting in 41 girls (ages 9-11) spread across two classes. Each design session lasted 45 minutes and occurred approximately twice a month throughout the year for a total of 18 sessions. In this paper, we will refer to this implementation as “Y2 Private.”
All Y1 and Y2 making sessions followed approximately the same overall structure shown in Table 1. Introduction sessions introduced participants to the tools and techniques they would use throughout the workshop. Following these introductory activities, builders interviewed their clients about their dream toys asking: “What kind of toys do you like? If you could imagine any toy, what would it look like and how would you play with it?” The next sessions included opportunities to brainstorm possible toy designs that would meet the requests of the clients, prototyping, and eventually building the final toy. Finally, builders delivered their newly constructed toys to their clients during a “play date.” Two important changes to the BfT implementation in Y2 included allowing builders to revise their target client in Y2 Public and the addition of a client feedback and reflection session in Y2 Private.

Table 1: While the overall structure of each BfT implementation did not change, additional opportunities to interact with the clients and reflect on design ideas was added in year two

<table>
<thead>
<tr>
<th>Y1 Public (3 hr sessions)</th>
<th>Y2 Public (1.5 hr sessions)</th>
<th>Y2 Private (45 min sessions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>Introduction</td>
<td>Introduction</td>
</tr>
<tr>
<td>Omnianimal (wood)</td>
<td>3D card</td>
<td>Omnianimal (cardboard) x 3</td>
</tr>
<tr>
<td>Client Interview (20 min)</td>
<td>Client Interview (20 min)</td>
<td>Client Interview (20 min)</td>
</tr>
<tr>
<td>Brainstorm</td>
<td>Collage of Client Interests</td>
<td>Brainstorm x 2</td>
</tr>
<tr>
<td></td>
<td>Revise Client</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brainstorm</td>
<td></td>
</tr>
<tr>
<td>Prototype</td>
<td>Prototype x 2</td>
<td>Prototype x 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Client Feedback</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design Revision</td>
</tr>
<tr>
<td>Build Final Toy</td>
<td>Build Final Toy x 3</td>
<td>Build Final Toy x 7</td>
</tr>
<tr>
<td>Play date (30 min)</td>
<td>Play date (30 min)</td>
<td>Play date (30 min)</td>
</tr>
</tbody>
</table>

Each session was video recorded and detailed field notes were taken by researchers present in the room. The research team conducted semi-structured interviews with participants to determine makers’ prior experiences with technology, construction, and crafts as well as knowledge of relevant concepts or skills. In Y1 Public, most all participants were interviewed both before and after the workshop. In Y2 Public, all participants were only interviewed before the workshop due to scheduling difficulties. Twelve participants were randomly selected to be interviewed before and after the BfT implementation in Y2 Private. Interviews were video recorded and transcribed. Transcripts were coded “bottom-up” where coding categories emerged from patterns in the data. A subset of the codes relevant to this analysis were applied by outside coders. Interrater reliability was computed for each data set with all sets achieving greater than 0.7 Cohen’s Kappa initially and improving to greater than 0.9 after discussion.

A large number of artifacts were produced throughout this project. These include worksheets (My Client Profile, client feedback, etc.) and notebooks completed by the builders, photographs and roaming video recordings of participants working on their projects, and photographs of toy designs throughout the construction process. These artifacts provided a broad picture of each participant’s work, such as whether they worked independently or alone, as well as their level of expertise both in technique and constructed toy.

**Results**

The central hypothesis of the BfT program was that building for others would be a compelling way of framing the value of making for young girls. We expected that young girls might be excited to work with younger children in their community, but also assumed that the quality of the builder-client relationship would impact the degree to which this activity felt meaningful to the builders. Each implementation provided a different builder-client relationship which allowed us to interrogate the role this relationship played throughout the construction process.

**Bots for Tots year one**
In the first implementation of the BfT project, we invited 4th grade builders to create toys for pre-Kindergarten (pre-K) children (age 4) in their school. Because the number of students in the pre-K class was greater than that of the builders in the BfT project, Y1 Public builders were broken into groups of two and paired with five to six pre-K clients. Consequently, when the builders first interviewed their clients and asked them to describe their dream toys, each pre-K child described a different toy resulting in a large number of highly diverse requests. For example, one set of requests included a Pinky Pie pony, a phone, and a monster truck race car.

Rather than abandon some ideas in favor of a more coherent design, teams took seriously each request by the children. Some groups addressed this difficulty by combining many ideas into one toy, such as Tayla and Inez’s “shopkins” plane car. Others chose to create multiple objects that could work together to satisfy the design requests. For example, Kelly, Raquel, and Kyle created a toy that included a Pinkie Pie pony doll phone that rides a monster-truck-skateboard (Figure 1).

Figure 1. Groups took the many diverse requests by the clients seriously, going as far as to merge a variety of ideas into multiple toys that worked together, such as the Pinkie Pie doll phone that rides a monster truck skateboard.

The desires and requests of the clients also played a role in early phases of the construction process. When making design decisions about toys, teams often took on difficult tasks to meet the requests of the pre-K client. For example, when trying to decide which materials to use when making the wings and wheels of a toy plane, a builder from another team suggested something “squishy” so that the toy would be pleasant to cuddle. Another designer quickly spoke up stating, “I disagree that the wings should be squishy! Cause isn’t it an airplane and is supposed to fly? So how’s it going to fly if the wings are squishy!”

When eight Y1 Public builders were interviewed after the project (one of the girls was unable to schedule a post-interview), three builders expressed some disappointment in what they created and six indicated they would like to add to or modify their designs. However, all eight interviewed children told us they felt their design was a success. For example, Juan stated, “I felt happy because the kids were happy and they were like each one were playing with the toys!” Similarly, when asked if they’d rather make for themselves next time or for someone else, five of the six girls interviewed told us they’d prefer to make for someone else (due to an error made by the interviewer, one of the girls was not asked the question during the interview, though responding to what made her proudest, she stated she liked seeing the kids’ faces, “it was so nice!”).

Bots for Tots year two

In the second year of implementations we sought to increase the personal nature of the client-builder relationship. Y2 Private offered a unique opportunity to leverage an existing mentorship program that had the 4th grade girls of the school mentoring 1st graders (age 6) throughout the year. Using this existing program, our BfT implementation in this school paired each 4th grade “Big Sister” 1:1 with her 1st grade “Little Sister” client (because of uneven numbers, one client had two builders). The builder-client pairs interacted multiple times throughout the school year (in addition to those facilitated by the BfT project) which allowed each builder to get to know the interests and personality of her client. This relationship was central to how and what the Y2 Private builders created.

Early in the semester builders interviewed their clients asking them to describe a dream toy. Using a “client profile sheet” builders were encouraged to identify the interest and likes of the client—including toys, materials, colors, etc.—as well as any objects or things they do not like. In a few cases, this interview provided builders with a precise specification for a dream toy to be built—for example Ava’s client indicated she wanted a “pop out monster car.” In most cases, however, this interview provided a list of potential toy features or possible toy properties that could become part of a toy design. For example, Chloe’s client profile sheet suggested her client likes stuffed animals, bounce balls, Magna-Tiles, and drawing, and does not like sharks.
Using these lists, builders brainstormed toy ideas that might satisfy the requests of their clients and drafted lists of materials and tools they might need to create this new toy. For example, Chloe starred “stuffed animal” and “draw” on her list of toys and activities provided by her client. She then brainstormed ideas that included extensive details about the materials and colors she would use in her construction. While Ava knew she was to build a “pop out monster car,” a variety of concerns would need to be considered. For example, Ava chose to make the car out of wood, planned to include a “squishy guy” with a “funny wig,” and intended to have the car play music and light up when used. Ava’s description also included a constraint expressed by her client, “No pink or purple on or in the car.”

Brainstormed ideas were eventually translated into a simple prototype which was then shown to the client to gauge her interest in the design and to received feedback and suggestions for the final dream toy. Occasionally, the client simply provided additional ideas such as “more eyes” for Ava’s pop up monster car and “fluffier” for Chloe’s whiteboard bunny. However, others requested large changes. For example, when Ruth presented her client with a prototype of a pillow with a voice recorder that spoke “I love you!” when squeezed, her client told her she would prefer a doll house. Ruth was initially disappointed, but after some encouragement from her classmates, she was soon back on track sawing sheets of plywood to build the new toy requested by her client.

Incorporating this feedback into their designs, the 4th grade builders then constructed the final toy over a few weeks and eventually delivered these toys to their clients in a final “play date” at the end of the year. Of the 40 projects being built by Y2 Private builders, 26 of them stayed consistent from the brainstorming phase all the way to the final toy construction. 14 of the builders made changes after the initial brainstorm phase or after receiving feedback on their prototype from the client. The reasons for these changes varied. Three of the 14 altered their design because they became enamored with a different material or because the materials available did not meet the requirements of their designs. Two of these girls changed their design due to requests by the client. For example, Sari had originally planned to make a stuffed talking monkey as requested by her client. However, when playing with the prototype her client informed her she would rather have a white talking sheep. Sari used this feedback to alter her final toy design and produced a beautiful white fluffy sheep (Figure 2).

![Figure 2](image_url)

Many girls altered their designs between the brainstorm, prototype, and final construction phases—such as Sari’s shift from a monkey to a sheep design seen here—due to requests made by the client.

Eight of the 14 that altered their designs between the brainstorming and final toy stage did so because they had difficulty achieving the goals they set out in their design. These challenges varied from sawing wood the correct length to cutting fabric for different shapes for stuffed animals. All eight of these girls instead chose to make a pillow for their clients (an activity they had done in class the previous year). While the pillow design was not the design originally requested by the clients, all eight designs were personalized for each client and incorporated features that were requested, such as using the client’s favorite color, sewing the client’s name into the pillow, or drawing pictures that were of interest to the client.

After the year-long project, 12 randomly selected Y2 Private participants were interviewed. Similar to Y1 Public, we asked participants if in a future BtT project they would prefer to build for themselves or for someone else. Ten of the 12 girls said they would like to make for others. LillyJane told us, “I like making stuff for others because I love seeing their reactions.” Panni explained, “Well, they really appreciate what I make for them and I love hugs and they give me hugs!” Five of these 10 girls expressed interest in making something for themselves as well as others. Two told us they’d prefer to just make something for themselves next time. These girls felt proud of the toy they had made and were sad to part with their construction.

Y2 Public offered a different set of results regarding the role of the client in the builders’ designs and construction processes. As with the other conditions, we had planned for builders to create dream toys for
younger children in their school. Similar to Y1 Public, we created teams of two builders and provided an opportunity for them to interview 5-6 pre-K clients. As builders worked to document their clients’ likes/dislikes in the construction of a collage, it became clear that they had little interest in building the toys requested by the pre-K children. One boy felt the “squeaky gem toy” requested by his clients was boring, others seemed more interested in making a collage using pictures relevant to the then upcoming 2016 presidential election.

As the goal of the project was to leverage builder’s goals and values in the making process, we allowed builders to choose new clients. Because anxiety was high for children in this school—which includes many immigrant families—after the 2016 presidential election, we suggested builders might create toys for people they were close to as a way to show they cared. Builders then generated a new list of potential clients. All builders included their friends (five of which included other people in the class) on their potential client list. Two also listed their mothers and two listed the pre-K clients they had interviewed previously.

After choosing new clients, Y2 Public builders became reenergized about the construction process. One of the groups continued to build for their pre-K clients, creating and sewing a stuffed T-Rex monster that met the original requests of the pre-K clients they had interviewed. Another builder chose to work alone to create a toy train for one specific pre-K student that was the younger brother of a fellow BfT builder. Two other groups and a third solo builder chose to make toys for their friends.

**Discussion**

The BfT project sought to examine the impact of leveraging “ways of knowing” to increase young women’s participation in construction activities. The results above highlight how the changes in implementation design as well as the school and wider communal context impacted the young makers’ interests in building for others. While the idea of making for others was compelling for most girls (and some boys) in all three implementations, the degree to which this framing impacted practice and future interest varied by context and design.

Y1 Public provided an existence proof for the notion that building for others might be a compelling way to engage young girls in construction workshops, and the Y2 implementations allowed us to make several modifications to the design of the project to further interrogate key features of the project system. In particular, Y2 Private allowed us to examine a one:one builder-client ratio and in Y2 Public we replicated the few:many builder-client ratio from Y1. We find that the quality of the relationship between builders and clients matters a great deal. Specifically, while nearly all young girls found making for others compelling, when able to make for one specific person, rather than many at once, builders more frequently considered the client’s interests in all phases of the design process.

In Y1 Public, teams of two to three builders took on the task of creating a dream toy for five to six pre-K children. While teams rarely mentioned individual pre-K children, during the brainstorm phase they took each request by their pre-K clients seriously. As indicated in the results section, this meant that teams often created many toys that worked together, or one toy that blended each idea into one artifact. However, this attention to the client took place almost exclusively in the brainstorm phase. Once teams identified a project idea, the client was only occasionally mentioned. When the client was mentioned, it was often to motivate team members that were not contributing equally.

While the results from Y1 Public were compelling, in Y2 we intentionally set out to better support the relationship between builders and clients to reflect Belenky et al. (1986) connected knowing—women’s tendency to value social interactions and a sense of community. Leveraging the Big/Little Sisters program ensured builders would not only encounter their client more frequently (they shared the same building and often met outside of class time), but also be encouraged to develop a mentor-like relationship providing care and guidance to their client as it was their first year of elementary school. This intimacy seemed to motivated the builders throughout the making process. For example, Hailey said in the post interview that she “really wanted to make something that [her client] would remember and really love.” Kate told us she sometimes asked her clients during recess if she still wants a doll to make sure that she was making it right for her.

In all three iterations builders had a high degree of flexibility in what they made. While they were encouraged to make the client’s dream toy, there was no metric or standard to measure their adherences to this program goal. However, all builders, made some effort to do so. In Y1 Public, builders took on challenging and even occasionally impossible design tasks. The one group that struggled to complete their design on time, reconceptualized their project as a DIY for their clients creating an opportunity for the builders and clients to build together during the play date (Figure 3). In Y2 Private, client feedback and reaction had both positive and negative impact on the builders and drove their design decisions. As shown in the results section, Chloe’s client liked her whiteboard bunny so Chloe only made minor changes to her design; while Ruth changed her project entirely when her client expressed dislike for her pillow prototype. Out of 40 Y2 Private projects, 39 were personalized to their clients. In Y2 Public the lack of interest in the initially assigned pre-K clients caused some
builders to lose interest in the project all together. After allowing participants to choose new clients, each team became motivated to persist through the design process. While we did not see the same degree of commitment to the client as that of Y2 Private, nor did we see the complex designs of Y1 Public, designers did produce complete and compelling artifacts aligned with what they believed to be the interests of their clients (Figure 3).

Finally, while most of the girls found building for younger children in their school community compelling, some did not. The two girls in Y2 Private that suggested they would prefer to make for themselves next time indicated that they were sad to part with a project for which they had worked so hard. One girls explained by stating, “I mean once you make something, I kind of like to keep it so I can like admire it.” Likewise, many of the boys did not find the pre-K clients compelling. One of the Y1 Public boys told us he would like to make a toy for himself. The other boy in this implementation indicated he would like to make a toy for his younger brother rather than the pre-K children he did not know very well. Two of the Y2 Public boys happily made toys for the pre-K children, while the other four preferred to make for a friend or family member.

In each of the three implementations, making was explicitly framed as a way of giving to others. However, the reason for giving depended on more than the gender of the builder. In Y1 Public, builders created toys as a gift to younger members of their community. The relationship between builder and client was modest, but effective at driving initial construction activity. This relationship was enhanced by leveraging the mentor-mentee relationship in Y2 Private. For these girls, building a dream toy became a way of representing and strengthening this mentor-mentee relationship. While we assumed our activity framing from Y1 Public would work similarly in Y2 Public, we instead found external social and political pressures meant a narrowing of what counts as one’s community. For Y2 Public builders, existing personal relationships were paramount, and making was a way of reinforcing that relationship.

These results suggest that while framing making as a way of building connections to one’s community will likely be compelling for young makers—and girls in particular—the degree to which this framing drives the making process will depend on who builders see as being part of their community and the quality of that connection. We suggest designers of maker activities should leverage “making for others” as a way of increasing participation and persistence among girls in maker activities, while also acknowledging the challenge of determining a compelling “other.” In addition to using the existing research literature, as we did for Y1 Public, we recommend designers seek to both connect with existing community programs as well as build into the activity design opportunities for participants to express with whom they desire to connect.

Conclusions
Women have been makers for centuries. And yet, the recent conceptualization of “making” elevated by the maker movement and pushed by tech companies, startup culture, and schools alike often legitimizes masculine construction activities and materials over those traditionally performed by women. As these maker activities have become a primary way of engaging young learners in STEM domains, this new definition of making must be challenged.

One method to address this overtly masculine definition of STEM has been to design maker and STEM workshops that focus use traditionally feminine materials and interests. We propose to further broaden this effort by recentering women’s ways of knowing in maker practices. In our work, we have found that when making is framed as being a way of building connections to one’s community, young girls are likely to see these activities as aligning with their values and goals.
Of course, ways of knowing and connecting will vary along a variety of dimensions. For some this connection may be strongest when positioned in a mentorship role—as we saw in our own Y2 Private implementation—while for others, building for family or a friend may be most compelling. But making and construction isn’t only an interaction between builder and tools or materials. Both bring with them cultural expectations, goals, and values. Future efforts to broaden participation in STEM domains through making and construction must acknowledge and center the values and goals of diverse learners in both design and implementation.

References


A Study of Collaborative Knowledge Construction in STEM via Educational Robotics

Chrysanthos Socratous and Andri Ioannou,
chrysanthos.socratous@cyprusinteractionlab.com, andri.i.ioannou@cut.ac.cy
Cyprus Interaction Lab, Cyprus University of Technology

Abstract: The educational robotics learning environment can integrate the benefits of robotics technology, computer supported collaborative learning (CSCL) and problem-based pedagogy, in an authentic learning space, simulating real-world problems. This study investigated primary school students’ patterns of knowledge construction in STEM, as they engaged in collaborative problem-solving using educational robotics. Data analysis involved micro-level examination of students’ discourse and interactions with their peers, the teacher and the robot, and students’ delivered solutions (software programs and worksheets). The study presents three conditions that appear to relate to higher levels of knowledge construction: (i) embodied interaction with the robot; (ii) fair contribution by teammates adhering to predefined roles, and (iii) cognitive dissonance as a result of the robot’s failure to perform the expected outcome. The study contributes to the design of educational robotics learning environments and conditions for collaborative knowledge construction in the STEM field.

Introduction

Since the beginning of the 21st century, there is a growing discussion about changes that should be made in the education systems in order to be able to meet the needs of the borderless, globalized and internationalized world. Problem-solving, critical thinking and collaboration are typically identified as key skills for success in the 21st century. Many reform efforts of the educational systems of advanced countries aim towards this direction. Frequently, learning in the science, technology, engineering, and mathematics (STEM) domains are cited as vehicles for the development of these skills for students (Ardito, Mosley & Scollins 2014).

Contemporary educational inquiry highlights the need to incorporate a technology-rich learning environment into the teaching of the STEM domains. The technology offers support for learning STEM-related school subjects or for enacting cross-curricular learning activities. Among flourishing arrays of technologies, educational robotics could be a highly effective learning environment for teaching STEM concepts (Barker et al., 2008). Educational robotics provides the opportunity for real-world application of science concepts, technology engineering, and mathematics, and helps to remove the abstractness of these fields (Nugent et al., 2010).

In this work, we present a learning experience using educational robotics, allowing students to engage in collaboration and co-construction of shared understandings in the fields of STEM. Our goal was to: a) field-test our design and explore patterns of collaborative knowledge construction, and b) explore the mediating role of educational robotics in our context. Below, we first present past papers on educational robotics and collaborative knowledge construction. Next, we present the theoretical grounding for our design and we detail the specifics of our activities and implementation procedures in a classroom setting. We conclude the section by presenting our findings from a field study with 14 student-participants and discussing possible links between the use of educational robotics and the observed outcomes.

Background work

Educational robotics

Our review of the literature reveals that educational robotics are used in education as a tool for achieving the following main objectives: a) to teach STEM, b) to develop learning skills such as problem solving, collaboration, scientific inquiry and critical thinking and c) to foster student motivation to engage in science and technology (Nugent et al. 2010). According to Eguchi (2010) most of the empirical work in education has focused on the first main objective. The majority of this work explores topics related to the field of physics and mathematics and emphasizes on skills that can be developed through robotics such as, problem solving skills, logic and scientific inquiry (Benitti, 2012). However, few studies have dealt with how knowledge is constructed and what elements of educational robotics can be associated with or can promote collaborative knowledge construction.
Many studies of educational robotics have been conducted in the recent years. These investigations have focused mainly on higher education; fewer studies have been conducted with secondary and elementary school students. Educational robotics with K-12 students have been implemented mainly out of schools, in summer or after-school programs and competitions (Barker et al., 2010). In general, there is limited empirical evidence to prove the direct impact of educational robotics on the K-12 curriculum. Many educators, researchers and educational theorists (e.g., Papert) believe that robotics can provide an enormous source of energy that can be used to stimulate children’s learning. Overall however, while the arguments for the benefits of educational robotics in the classroom continue to grow, scientific inquiry in the area is sporadic and most of the literature is descriptive in nature. The processes and conditions under which any specific learning goals are achieved are far from being documented. Thus we need first to explore what educational robotics have to offer to education, before we arrive in overestimated and possibly wrong conclusions regarding their pedagogical value. Another main weakness in the area of educational robotics is the absence of clearly defined curriculum and educational material for teachers. Alimisis (2013) pointed out that there is no systematic introduction of robotics in school curricula within the European school systems.

Educational robotics and STEM-related outcomes
There are already a number of reports from robotics programs claiming that educational robotics can improve the performance of students in STEM (Bers & Portsmore, 2005; Nagehaudhuri, Singh, Kaur & George, 2002, as stated in Bilotta et al., 2009). For example, the use of educational robotics led to the enhancement of students (a) mathematics performance (Highfield, 2010; Nugent et al., 2009) (b) science performance and physics content knowledge (Mitnik et al., 2008) (c) engineering design skills (Larkins, Moore, Rubbo & Covington, 2013; Hong et al., 2011), and (d) STEM knowledge (Barker et al., 2010; Barker & Ansorge, 2007). Other benefits from the use of robotics include improvement in collaboration skills (Nugent et al., 2009; Mitnik et al., 2009) problem-solving skills (Nugent et al., 2009), creative thinking (Barak & Zadok, 2009), and scientific inquiry (Sullivan, 2008).

Educational robotics and collaborative knowledge construction
In the last 20 years, researchers have seriously studied learning in small groups and the nature of cooperation and interaction have turned into a focal issue for research on learning in social settings. Essential to collaborative learning is knowledge construction where the collaborative learning aims at co-constructing knowledge upon sharing information in groups for solving given tasks (Alavi & Dufner, 2005). In the recent years, the focus of knowledge construction moved from knowledge attainment to skill development knowledge in order to prepare students for the challenges of the 21st century (Wen, Zaid & Harun, 2015). The focus moved from simply gathering information to a more complex process of researching and thinking critically about the new information in order to use it in meaningful ways.

The joint construction of knowledge allows learners to experience a greater level of understanding (Kafai & Resnick, 1996) because they must construct their own knowledge to learn the truth (Tam, 2000). Knowledge is constructed by students when they participate and evaluate their own learning. Collaborative knowledge construction encourages students to investigate deeper about a subject so that can reach their highest potential level of development. The development of new understanding is coming as a combination of prior knowledge and skills with new experiences.

An educational robotics class can potentially contribute to the collaborative knowledge construction process. In a learning environment, educational robotics has the role of mindtools. The term “mindtools”, as proposed by Jonassen (2000) in the sense of cognitive tools, represent the constructionism dimension of constructivism. Using educational robotics as mindtools, in a classroom, we apply constructivism -- students construct a physical object, while at the same time they construct problem solving knowledge. Learning is no longer teacher-centered but knowledge is actively constructed by the learner (Harel & Papert, 1991). Students can change or negotiate their existing knowledge into explicit knowledge. Knowledge construction is therefore, formed through a dual pathway; through interaction with the artifact and through interaction with peers. Several studies indicate that educational robotics can be used as mindtools supporting knowledge construction through the design of meaningful artifacts in authentic projects, learning by doing, facing cognitive conflicts and learning by reflection and collaboration (Mikropoulos & Bellou, 2013; Jonassen, 2000).

Learning design
Pedagogical aspects
Our work draws on social constructivism, which emphasizes the importance of social interaction in knowledge construction (Palincsar, 1998). Many popular educational formats such as problem-based learning and computer-supported collaborative learning (CSCL) have their roots in social constructivism, which states that group discussions can help students learn. In this work, we see educational robotics as fully compatible with the nature of collocated CSCL. The technology provides a way to infuse real world experiences to the CSCL setting, through the hands-on nature of collaborative activities they can facilitate. Also, in line with problem-based learning pedagogy, we aim for students, in groups, to engage in problem-solving processes such as defining the problem, developing a strategy, testing and experimenting, and reflecting in and on action (Ioannou, Vasiliou, Zaphiris, 2016). Problem-based learning helps situate learning in meaningful tasks and emphasises the importance of practical experience in learning (Hmelo-Silver, 2004; Ioannou, Brown, & Artino, 2015). In CSCL environments, participants are actively engaged in creating or co-constructing knowledge. Although in most CSCL studies, collaboration takes place through a computer network, face-to-face interactions have proven to be the richest communication media, conveying the greatest social presence (Newberry, 2001), and consequently, increasing the quality of learning and the achievement of learning objectives (Aspden & Helm, 2004).

Educational robotics activities and processes
Designing the technology-enhanced learning experience was a task undertaken by a teacher and an educational technologist. We used the Lego Mindstorms EV3 toolkit. Content for the activities came from: a) the national curriculum on mathematics and science education, and b) the EV3 STEM curriculum. Typically, the teacher presented students with a challenge and a mat. There was no clear path to the solution; students could adopt any strategy to come to a solution to the challenge. The teacher acted as a facilitator, supporting student’s thinking without providing any answers. Upon completion of each task, a debriefing phase took place: groups demonstrated their strategies in addressing the challenge and they answered questions asked by the teacher and students in other groups. The teacher facilitated discussion on best strategies and reflection on what kinds of problem-solving and STEM skills were learnt. Sample activities are presented in Table 1 while Figure 1 summarizes the learning cycle.
measurer interacts with the recorder to report data for recording. The recorder interacts with programmer to indicate the readiness of the data collection for further action or give auxiliary data based on current data collection that may help with programming the robot. Roles were randomly assigned to students. From task to task the teacher ensured that the roles rotated amongst the team-members.

Table 1: Sample educational robotics learning activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Explanation</th>
<th>Main STEM Pillars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maze challenge (80 min)</td>
<td>Groups program the robot to move from its starting position through a without touching any walls.</td>
<td>Numbers &amp; calculations, robot sensors, robot wheels diameter and speed vs. turns, loops, measurements</td>
</tr>
<tr>
<td>Robot-slalom challenge (80 min)</td>
<td>Groups program a robot to move along the outside of each flag and cross the finish.</td>
<td>Numbers &amp; calculations, robot sensors, geometrical symmetry, swing turn and point turn, loops, measurements</td>
</tr>
<tr>
<td>Draw a hexagon challenge (80 min)</td>
<td>Students program their robot to draw a hexagon using a gyro sensor.</td>
<td>Numbers &amp; calculations, polygons, supplementary &amp; complementary angles, internal &amp; external angles, design a pen holder, measurements</td>
</tr>
</tbody>
</table>

Field study

Participants, setting and data collection

Our sample was composed of 14 students (6 males) in Grades 4, 5 and 6 (9-11 years old) who attended a public elementary school in Cyprus. Students were divided into 4 groups, mixing gender, technological, and problem-solving abilities. There were two weeks of preparation activities to help students get familiar with the EV3 robot (e.g., move straight ahead, turn base on some angle, use sensors, robot decisions e.g., loops), followed by three 80-minutes sessions of STEM problem-solving activities. Two cameras were placed in the room to fully cover student interaction and technology use. Verbal contributions were captured separately via audio recorders next to each team; audio was later synced with the video.

Coding of video data and plotting chronological diagrams

All video data were transcribed verbatim and content analyzed. We used the coding scheme reported in Gunawardena, Lowe, & Anderson’s (1997) Interaction Analysis Model, which conceptualizes the level of social and collaborative knowledge construction. The analysis focused on the “unit of meaning”, each unit fitting into a sole category of the coding scheme (e.g., from minute 0:30 to 0:35 teammates share ideas on potential strategies). Overlapping talk was attributed to the most dominant category and team member. Around 30% of the video was coded by the first researcher, with a second researcher independently coding the same units. Reliability was high (agreement over 90%) and therefore, the first researcher finished coding the complete dataset. Table 2 presents examples of application of the coding scheme.

Table 2: Examples of application of the coding scheme of collaborative knowledge construction

<table>
<thead>
<tr>
<th>Level</th>
<th>Operation</th>
<th>Example excerpts from the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>KC-1: Sharing/ comparing/ adding of information.</td>
<td>Statement of observation, opinion or a background information; Definition description or identification of a problem;</td>
<td>P2: “We will use the ultrasonic sensor to avoid the flags.</td>
</tr>
</tbody>
</table>
For a chronological investigation of within-group collaboration, we further plotted student’s discourse and activity on chronological diagrams. We used the CORDTRA technique, initially suggested by Hmelo-Silver et al. (2011). This technique of visual representation enables one to combine the chronological visual of discourse with other types of coded data allowing for the examination of patterns and behavioral sequences.

Findings and discussion

Embodied interaction with the physical robot is linked to higher levels of knowledge construction

Based on Table 3, out of 172 coded units, the majority of verbal interactions (38.4%) was coded in the lowest level of knowledge construction (KC-1). This was followed by students’ KC-4 level involving 43 verbal units (25%) and KC-3 level involving 37 verbal units (21.5%). KC-2, appeared with a relatively lower percentage than expected with only 18 coded units or 10.6%. The highest level of knowledge construction (KC-5), was difficult to achieve and was only represented by 8 units (4.5%). Gunawardena, Lowe, and Anderson’s Interaction Analysis Model has been almost exclusively used only in online learning discourse in CMS and CSCL settings (e.g., Ioannou, Demetriou & Mama, 2014). According to these studies, KC-1 statements accounted for the largest percentage of the overall discussion and were prerequisite for subsequent higher levels of knowledge construction. The findings of the present study differ from typical results of online learning activity in that, KC-4 accounted for the second largest proportion of discourse units.

Table 3: Number of Codes Across Levels of Knowledge Construction and Groups

<table>
<thead>
<tr>
<th>Level of Knowledge Construction</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharing/adding (KC-1)</td>
<td>20</td>
<td>15</td>
<td>12</td>
<td>19</td>
<td>66 (38.4%)</td>
</tr>
<tr>
<td>Exploration of dissonance (KC-2)</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>18 (10.6%)</td>
</tr>
<tr>
<td>Negotiating meaning (KC-3)</td>
<td>16</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>37 (21.5%)</td>
</tr>
<tr>
<td>Testing synthesis (KC-4)</td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>15</td>
<td>43 (25%)</td>
</tr>
<tr>
<td>Applying co-constructed knowledge (KC-5)</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>8 (4.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>41</td>
<td>25</td>
<td>49</td>
<td>172 (100%)</td>
</tr>
</tbody>
</table>

The increased KC-4, compared to previous CMC and CSCL studies, lead us to hypothesize that educational robotics might have encouraged knowledge construction at this level because of the hands-on experimentation and embodied interaction with the physical robot. We therefore, used chronological diagrams, to pinpoint what students were doing, when they exhibited KC-4 of knowledge construction. By zooming into the groups’ chronological diagrams, we found that “Execution of Plan” was tightly coupled with higher levels of knowledge construction. That is when students interacted with the physical robot to execute their plan, they

P3: “No, we tried using the ultrasonic sensor when we solved the maze challenge and took us a lot of time. What do you think”?

P4: “I will draw a hexagon to find how many triangles are formed.” P5: “So, 4 triangles multiplies by 180° equals 720° divides by 6 angles...(thinking) 120°”.

P5: “No, what is he doing? It is turning too much (while observing the drawing of the robot). What went wrong?” P6: “There is a problem with the gyro sensor. Let’s remove it and put it back”

P13: “Yes, that’s it. Bravo! The robot must turn as much as the supplementary angle of the internal angle. That was clever.”

For a chronological investigation of within-group collaboration, we further plotted student’s discourse and activity on chronological diagrams. We used the CORDTRA technique, initially suggested by Hmelo-Silver et al. (2011). This technique of visual representation enables one to combine the chronological visual of discourse with other types of coded data allowing for the examination of patterns and behavioral sequences.

Table 3: Number of Codes Across Levels of Knowledge Construction and Groups

<table>
<thead>
<tr>
<th>Level of Knowledge Construction</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharing/adding (KC-1)</td>
<td>20</td>
<td>15</td>
<td>12</td>
<td>19</td>
<td>66 (38.4%)</td>
</tr>
<tr>
<td>Exploration of dissonance (KC-2)</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>18 (10.6%)</td>
</tr>
<tr>
<td>Negotiating meaning (KC-3)</td>
<td>16</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>37 (21.5%)</td>
</tr>
<tr>
<td>Testing synthesis (KC-4)</td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>15</td>
<td>43 (25%)</td>
</tr>
<tr>
<td>Applying co-constructed knowledge (KC-5)</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>8 (4.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>41</td>
<td>25</td>
<td>49</td>
<td>172 (100%)</td>
</tr>
</tbody>
</table>

The increased KC-4, compared to previous CMC and CSCL studies, lead us to hypothesize that educational robotics might have encouraged knowledge construction at this level because of the hands-on experimentation and embodied interaction with the physical robot. We therefore, used chronological diagrams, to pinpoint what students were doing, when they exhibited KC-4 of knowledge construction. By zooming into the groups’ chronological diagrams, we found that “Execution of Plan” was tightly coupled with higher levels of knowledge construction. That is when students interacted with the physical robot to execute their plan, they

P3: “No, we tried using the ultrasonic sensor when we solved the maze challenge and took us a lot of time. What do you think”?

P4: “I will draw a hexagon to find how many triangles are formed.” P5: “So, 4 triangles multiplies by 180° equals 720° divides by 6 angles...(thinking) 120°”.

P5: “No, what is he doing? It is turning too much (while observing the drawing of the robot). What went wrong?” P6: “There is a problem with the gyro sensor. Let’s remove it and put it back”

P13: “Yes, that’s it. Bravo! The robot must turn as much as the supplementary angle of the internal angle. That was clever.”
often engaged in KC3-KC5 (see Figure 2). Students were engaged in a process of “conversation with the robots”, through which they promoted self-directed learning and engaged in the construction of new knowledge. The physical and embodied interaction with the robot gave students the opportunity to test or modify their new synthesis (KC-4) against existing cognitive knowledge, personal experiences, and data.

**Fair contribution by teammates adhering to predefined roles is linked to higher levels of knowledge construction**

We found that collaborative knowledge construction was more evidenced in some groups than others. Groups 1, 2 and 4 appeared successful in engaging in the collaborative knowledge construction process, since their discourse involved contributions along all phases of knowledge construction. On the other hand, group 3 with only 25 coded units demonstrated limited engagement with the activities, whilst their discourse never reached the higher levels of knowledge construction. This case made us hypothesize that lack of within group interaction might have hindered collaborative knowledge construction. We therefore took a closer look at videos and chronological diagrams of all groups (see Figure 2; due to space limitations we present only the visuals of 2 groups) to pinpoint patterns of collaboration in relation to knowledge construction. We found that in groups 1,2 and 4 all teammates were active participants in the learning process, whilst they participated fairly, adhering to their predefined roles. Instead, members of group 3 did not serve their pre-defined roles and did not participate fairly in the tasks which seems to have led to failure in engaging in collaborative knowledge construction. It therefore appears that, assigning roles to teammates and serving these roles enabled fair contribution, individual accountability and social interdependence (Johnson et al., 1991) leading to better quality discourse and knowledge construction.

**Cognitive dissonance is linked to higher levels of knowledge construction**

A detailed examination of chronological visuals and associated groups’ discourse helped us understand the progressive interactions and breakdowns within each group. We noted that KC-2 (discovery and exploration of dissonance or inconsistency) was relatively rare in students’ contributions (only 10,6%). What’s more, when KC-2 type of discourse appeared, it took a while for the next level of contribution to appear. We therefore sought to understand cognitive dissonance and when it occurred. Zooming into the groups’ chronological diagrams and associated discourse, we found that cognitive dissonance was less often related to disagreement between the teammates and more often related to the robot’s failure to perform the expected outcomes during the execution of a planned strategy (stage 3 of Figure 1). In this case, students had to reconsider their strategies (i.e., going back to stage 2 of Figure 1). The finding suggests that the robot and its failure to deliver the expected result was a mediator to the discovery of cognitive dissonance or inconsistency; the latter was a time-consuming process which teammates struggled to overcome. Nevertheless, when the group overcame this stage, they engaged in higher levels of knowledge construction as evidenced in their chronological discourse. That is, inspection of the chronological diagrams of groups 1, 2 and 4 makes obvious that KC-2 contributions, i.e., cognitive dissonance mostly related to the robot’s failure to perform, are followed by contributions coded in the higher levels of knowledge construction (see Figure 2).
Conclusion
This work begins to collect the much-needed evidence around the practical utility and potential impact of educational robotics in school contexts. We described the design of a learning experience using EV3 educational robotics, allowing 14 students in the math and science classroom to engage in collaborative learning and problem-solving. We addressed the effectiveness of educational robotics in terms of collaborative knowledge construction in the STEM field. The findings demonstrated how knowledge is constructed and which elements of educational robotics and teamwork can promote collaborative knowledge construction in an educational learning environment. Overall, we offer a case study in which established pedagogy (problem-based pedagogy) blends with technological capabilities (robotics) to enable students to practically apply the key elements they are being exposed to in their STEM education curriculum. Educators can use these findings to develop interventions to assist students in engaging in higher levels of knowledge construction using educational robotics.

References


Collaboration Scripts Should Focus on Shared Models, Not on Drawings, to Help Students Translate Between Representations

Sally P.W. Wu, University of Wisconsin – Madison, pwwu@wisc.edu
Martina A. Rau, University of Wisconsin – Madison, marau@wisc.edu

Abstract: Students often struggle to translate between physical and virtual models when learning concepts in STEM courses. Collaborative activities may help students overcome these difficulties, especially if collaboration scripts prompt students to discuss shared representations. Which representation should collaboration scripts focus students’ interactions on? We investigate this question in a quasi-experiment with 560 undergraduate chemistry students. All students collaboratively built physical ball-and-stick models of molecules and translated them into wedge-dash drawings. Two experimental conditions received a collaboration script. For the model condition, the script prompted students to focus on the physical ball-and-stick models. For the draw condition, the script prompted students to generate intermediary drawings on paper. Compared to a control condition with unscripted collaboration, the model condition showed higher learning gains and the draw condition showed lower learning gains—especially for students with low spatial skills. Our results yield theoretical and practical implications for collaborative practices with multiple representations.

Keywords: Multiple representations, collaboration scripts, physical and virtual models, chemistry, spatial skills

Introduction

Figure 1. Physical ball-and-stick model (A) and wedge-dash structure (B). Each shows two isomers: molecules with the same molecular formula but different 3D arrangement of the atoms.

Figure 2. Students collaboratively work with physical ball-and-stick models and virtual wedge-dash structures.

Students in science, technology, engineering, and mathematics (STEM) often use multiple visual representations to illustrate abstract concepts (Ainsworth, 2008). For example, when learning about molecular geometry, students are typically asked to translate 3D physical ball-and-stick models (Figure 1A) into 2D wedge-dash structures (Figure 1B). To make these translations, they may work collaboratively with physical models and virtual representations (Figure 2). However, prior research shows that students often fail to spontaneously engage in effective collaboration strategies (Lou, Abrami, & D’Apollonia, 2001). Instructional support can alleviate these difficulties, for instance by providing collaboration scripts that prompt students to discuss how to translate between the representations. In our prior research, we found that such a collaboration script enhanced students’ learning compared to unscripted collaboration during the same activity (Rau, Bowman, & Moore, 2017).
Collaboration scripts for translating between multiple representations may be effective because they focus students’ collaborative interactions on shared representations. In our prior research, we observed two practices that are commonly promoted by instructors when students collaboratively translate physical into virtual representations. First, instructors often prompted students to focus on physical models they collaboratively constructed and to compare models directly to the virtual representation. Second, instructors often asked students to translate the physical representation into an intermediate drawing and compare the drawing to the virtual representation. Which of these instructional practices should collaboration scripts promote? We address this question in a quasi-experiment in which undergraduate chemistry students collaboratively translated physical ball-and-stick models into virtual wedge-dash structures to make sense of chemistry concepts related to molecular geometry.

Prior research
In the following, we briefly review research on students’ difficulties in translating among multiple representations and collaborating effectively. We then discuss how collaboration scripts may alleviate students’ difficulties by fostering the disciplinary practices of physical modeling and drawing.

Students’ difficulties in translating between multiple visual representations
Many STEM concepts are visuospatial and/or not directly observable. To make these concepts accessible, instruction in STEM domains uses multiple visual representations (Ainsworth, 2008; Rau, 2017). For example, chemistry instruction typically uses physical ball-and-stick models with wedge-dash structures (Figure 1) to illustrate concepts related to molecular geometry. Students benefit from multiple visual representations only if they can translate between representations to make sense of the underlying concepts (Rau, 2017). However, many students struggle with such translations (Ainsworth, 2008), which can impede their success in STEM domains, including chemistry (Stieff, 2007).

Translating between representations is particularly difficult for students with low spatial skills (Stieff, 2007) because translating requires students to map the visual features of one representation to corresponding features in the other representation (Rau, 2016). To do so, students must hold visual features in working memory and mentally rotate the features to align them (Hegarty & Waller, 2005). By definition, students with lower spatial skills experience higher cognitive load during this task, which can jeopardize their learning from multiple visual representations (Hegarty & Waller, 2005; Stieff, 2007). Indeed, the chemistry education literature documents that students with low spatial skills have more difficulties in translating between representations, which can result in lower learning gains (Stieff, 2007).

Students’ difficulties in collaborating with multiple visual representations
Collaboration can help students overcome difficulties in translating between representations because students can help each other map visual features and make sense of how representations show key concepts (van Dijk, Gijlers, & Weinberger, 2014). When working individually, students often fail to spontaneously reflect on their understanding of visual representations (Ainsworth, Bibby, & Wood, 2002). When working collaboratively, students may realize that they hold divergent views on the visual representations, which may prompt them to engage more deeply in making sense of the visual representations (Zhang & Linn, 2011). Further, collaboration may help students with low spatial skills receive support from peers to help them align and map features of different representations.

Yet, students often have difficulties in collaborating effectively (Lou et al., 2001; Weinberger, Stegmann, Fischer, & Mandl, 2007). Instructional support can help students overcome these difficulties, for instance by scripting their collaborative interactions. Such collaboration scripts can guide students’ interactions, for instance, by asking questions for students to discuss or by prompting them to explain concepts to one another (Fischer, Kollar, Stegmann, & Wecker, 2013; Weinberger et al., 2007). In a prior experiment (Rau et al., 2017), we tested the effectiveness of a collaboration script that adaptively provided prompts to discuss representations when students reached an impasse in translating between representations. Results showed higher learning gains for students who received the collaboration script than for students who worked on the same activities without the script.

Supporting collaborative practices with multiple visual representations
Yet, prior research on collaboration scripts has not examined how to focus students’ collaborative interactions on specific practices related to translating between multiple visual representations. Specifically, we observed two disciplinary practices that are common in classroom activities with multiple visual representations (NRC, 2012). Both of these practices serve to focus students’ attention on a shared representation.
First, instructors often prompt students to focus on shared physical models. Prior research suggests that interactions with physical models can help students learn domain knowledge (Stull, Hegarty, Dixon, & Stieff, 2012). For instance, students’ interactions with ball-and-stick models are constrained because the balls (atoms) have a designated number of holes for sticks (bonds), which are spread out as far apart as possible in a tetrahedral shape. Engaging in such interactions with the physical model can help students learn and retain information about the representation, such as how it shows molecular geometry and how to rotate the model for projection onto a 2D plane. Further, the ability to rotate physical models may also alleviate difficulties in mental rotation for students with low spatial skills (Barrett, Stull, Hsu, & Hegarty, 2014). However, much research suggests that models could hinder learning because students often do not know how to spatially align physical models with other representations to map features (Barrett et al., 2014). Because spatial alignment is particularly challenging for students with low spatial skills, focusing collaborative interactions on physical models could disadvantage students with low spatial skills.

Second, instructors often prompt students to draw additional visual representations on paper because it engages students in a valued disciplinary practice for STEM professionals (NRC, 2012). Professionals often draw intermediary representations to help them translate across representations during collaborations (Kozma & Russell, 2005). Prior research shows that drawing visual representations can help students learn domain knowledge (Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner, 2014; Zhang & Linn, 2011). Further, drawing activities may facilitate mental rotation because students can physically orient their drawings on paper to map visual features. Indeed, prior research shows that spatial skills do not affect students’ benefit from drawing activities (Schmeck et al., 2014). However, drawing activities without instructional support have been shown to result in high cognitive load, especially if the drawing task involves mental rotation (Schwamborn, Thillmann, Opfermann, & Leutner, 2011). This effect could disproportionately affect students with low spatial skills. Therefore, focusing students’ collaborative interactions on drawing may disadvantage students with low spatial skills.

In our prior study on adaptive collaboration script for translating between representations (Rau et al., 2017), we observed both practices. First, students used a shared modeling kit to construct physical ball-and-stick models. Consequently, when the collaboration script prompted them to discuss translation between the representations, they often focused on the models. Second, our classroom observations showed that students often drew wedge-dash structures on paper when they reached an impasse, a practice encouraged by the teaching assistants in the course. Hence, either of these practices could have accounted for the effectiveness of our collaboration script. However, our prior experiment did not control for whether or not students were prompted to engage in either type of collaborative practice.

Research questions
Our brief review of collaborative practices in STEM shows that it is, to date, unclear whether prompting students to focus on physical models or to generate intermediary drawings will best support their learning with multiple visual representations. Based on Learning Sciences theory, one can argue both for potential positive and negative effects of each collaborative practice. Further, both collaborative practices are common in STEM professions and often encouraged by instructors in STEM courses. Therefore, our goal is to investigate:

Research question 1: Is a collaboration script more effective if it focuses students’ collaborative interactions on physical models or on intermediary drawings in terms of enhancing students’ learning of domain knowledge?

Further, because translating between representations is particularly difficult for students with low spatial skills (Hegarty & Waller, 2005; Stieff, 2007), we investigate:

Research question 2: Do students’ spatial skills moderate the effects of model-focused and drawing-focused collaboration scripts?

Method
Chemistry course and participants
To address these questions, we conducted a quasi-experiment with 560 students in an undergraduate chemistry course at a Midwestern U.S. university. The course involved two 50-minute lectures attended by all students, one 50-minute discussion session, and one 3-hour lab session each week. Lab and discussion sessions were held in smaller sections of about 18 students each. The lab and discussion sessions were led by teaching assistants (TAs). All TAs received the same training in leading these sessions at the beginning of the semester. During the semester, students worked in small groups of 2-3 students during discussion and lab sessions.
Our quasi-experiment took place in a lab session which covered a topic related to molecular geometry: chemical isomers. Isomers are molecules made of the same atoms but differ in the spatial arrangement of their atoms. Instruction on isomers crucially relies on connecting the representations shown in Figure 1 because differences in the atoms’ spatial arrangements within molecules can have dramatic effects on the properties of chemical compounds. For example, if students fail to understand the difference between the left and right isomers in both Figure 1A and 1B, they may fail to understand that the melting point of the left isomer differs from the melting point of the right isomer.

Experimental design
The chemistry course had 34 lab sections. Two experimental conditions received a collaboration script. Specifically, five lab sections \((n = 78\) students) were assigned to the model condition that received a model-focused script. Six lab sections \((n = 110\) students) were assigned to the draw condition that received a drawing-focused script. The remaining 23 sections were assigned to a control condition \((n = 383\) students) that received no collaboration script. Students selected lab sections at the beginning of the semester to fit their schedule. We do not have any reason to believe that systematic differences exist between sections.

During the lab sections, all students worked collaboratively in dyads with the visual representations shown in Figure 1 on a sequence of chemistry problems. The problems required students to construct a physical ball-and-stick model of a specific isomer and to translate it into a wedge-dash structure. Further, each problem contained conceptual questions that required students to use the representations to make sense of concepts related to isomers.

Experimental conditions

The model and draw conditions received a collaboration script that prompted them to discuss how to translate between physical ball-and-stick models and virtual wedge-dash structures while they solved chemistry problems. As in our prior study (Rau et al., 2017), the collaboration script was implemented in an educational technology that presented the chemistry problems: Chem Tutor (see Figure 3). The Chem Tutor problems were created to be identical (i.e., same steps, same questions, same molecules) to problems on a paper worksheet traditionally used in this lab session. Chem Tutor instructed students to build the model collaboratively, input answers using a drop-down menu, and draw the wedge-dash structure in an interactive tool. If students made an error on a problem, the collaboration script provided immediate feedback by highlighting the part of the conceptual question or of the wedge-dash structure that was incorrect and prompting students to discuss it with their partners (Figure 3). Throughout all lab sessions for the model and draw conditions, the first author and several research assistants provided technical support with Chem Tutor. The TA for each session answered questions about the content, as they typically do for each lab session.

The difference between the model and the draw conditions regarded the introductory prompts in the collaboration script and an equivalent spoken prompt provided at the beginning of the lab session. Students in the model condition received prompts to “carefully build and orient their physical ball-and-stick models” before constructing virtual wedge-dash structures in Chem Tutor. Students in the draw condition received prompts to “plan their wedge-dash structures on paper” before constructing them in Chem Tutor. The respective prompts state that
the practice of “constructing models” or “drawing on paper” benefits students because it “aligns with the work of professional chemists and is an essential part of their reasoning process.”

**Control condition**

A “business-as-usual” control condition did not receive a collaboration script. Students solved the same sequence of chemistry problems using a paper worksheet traditionally used in this lab session. The worksheet contained the same problem-solving steps, conceptual questions, and isomers. Students also used a shared modeling kit to construct the physical ball-and-stick models, but they wrote down answers to the conceptual questions and drew wedge-dash structures on the worksheet without a collaboration script. At the end of the 3-hour lab session, TAs collected the worksheets to provide written feedback on problem solutions and wedge-dash drawings in the following week’s lab session.

**Assessments**

To assess students’ learning of domain knowledge, we used a pretest and posttest on isomers, evaluated in our prior study (Rau et al., 2017). A retention scale of the test assessed students’ ability to recall isomer concepts from the lab. A transfer scale assessed students’ ability to apply this knowledge to predict the stability of molecules.

To assess students’ spatial skills, we used the Vandenberg & Kuse test for mental rotation ability (Peters et al., 1995). This test was evaluated and used in prior research on chemistry learning (e.g., Stieff, 2007).

**Procedure**

We conducted our study as part of an undergraduate chemistry course. A lecture in week 3 of the semester covered molecular geometry and chemical isomerism. In week 4, students worked on activities in accordance with their typical lab schedule. First, as the required pre-lab exercise, students completed the pretest and spatial test online. Then, during their scheduled 3-hour lab session, students completed problems using Chem Tutor or worksheet that corresponds to their condition. Lastly, students completed the posttest online as the required post-lab exercise at the end of week 4.

**Results**

**Prior checks**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Spatial test Pretest</th>
<th>.541 (.186)</th>
<th>Retention scale Posttest</th>
<th>.628 (.210)</th>
<th>Transfer scale Pretest</th>
<th>.727 (.369)</th>
<th>Posttest</th>
<th>.764 (.375)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>.881 (.140)</td>
<td>.527 (.190)</td>
<td>.651 (.195)</td>
<td></td>
<td>.540 (.403)</td>
<td></td>
<td>.745 (.356)</td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>.875 (.132)</td>
<td>.515 (.170)</td>
<td>.630 (.202)</td>
<td></td>
<td>.654 (.397)</td>
<td></td>
<td>.801 (.326)</td>
<td></td>
</tr>
<tr>
<td>Draw</td>
<td>.857 (.166)</td>
<td>.541 (.186)</td>
<td>.628 (.210)</td>
<td></td>
<td>.727 (.369)</td>
<td></td>
<td>.764 (.375)</td>
<td></td>
</tr>
</tbody>
</table>

Because we used a quasi-experimental design in which students were not assigned to conditions at random but based on their lab section, we first checked for potential differences between conditions prior to the intervention. Table 1 shows the means and standard deviations of test scores by condition. A multivariate analysis of variance (MANOVA) with condition as independent factor and test scores (retention pretest, transfer pretest, and spatial skills test) as dependent measures showed no significant differences between conditions on the retention pretest ($F < 1$) or on the spatial skills test, $F(2, 570) = 1.24, p = .291$. However, there was a significant difference on the transfer pretest, $F(2, 570) = 10.61, p < .01$. Post-hoc comparisons showed that students in the draw condition had significantly higher scores than students in the control condition ($p < .01$). No other differences were significant at the pretest. To account for pretest differences, all following analyses use pretest scores as a covariate. Next, we checked for learning gains from pretest to posttest. A repeated-measures ANOVA showed significant learning gains on the retention test, $F(1, 570) = 145.63, p < .01$, as well as on the transfer test, $F(1, 570) = 75.37, p < .01$.

**Differences between conditions**

Because we conducted a quasi-experiment with assignment to condition by lab section rather than by student, we used a hierarchical linear model (HLM) to analyze differences between conditions. HLMs take into account nested sources of variance due to the fact that, for instance, students taught by the same teaching assistants tend to have more similar knowledge than students taught by different TAs. Specifically, we used an HLM that included a
random intercept for TAs. In addition, the HLM included pretest scores as a covariate to control for pretest differences prior to the intervention, condition as independent factor to test research question 1, and spatial skills and an interaction effect of condition with spatial skills to test research question 2.

On the retention posttest, there was no significant main effect of condition \( (F < 1) \) nor a significant interaction between condition and spatial skills \( (F < 1) \). On the transfer posttest, there was a significant main effect of condition on learning gains, \( F(2, 564) = 5.03, p < .01 \), such that the model condition outperformed the control condition and the control condition outperformed the draw condition. This effect was qualified by a significant interaction between condition and spatial skills, \( F(1, 564) = 4.75, p < .01 \). To gain insights into the nature of this interaction effect, we split students into groups with low spatial skills (0-33rd percentile on the spatial skills test), medium spatial skills (34th-66th percentile), and high spatial skills (67th-100th percentile). Post-hoc comparisons showed that the effect of condition was significant only among students with low spatial skills \( (p < .05) \), but not among students with medium or high spatial skills \( (ps > .10) \). Figure 4 illustrates these effects.

![Figure 4](image_url)  
**Figure 4.** Average scores on transfer posttest by condition and post-hoc splits into low (0-33rd percentile), medium (34th-66th percentile), and high (67th-100th percentile) spatial skills. Error bars show standard errors of the mean.

**Discussion**

This study investigated the effects of two disciplinary practices on a collaboration script that supports translation between multiple visual representations. Specifically, we tested whether a collaboration script is more effective if it focuses students’ collaborative interactions on physical models or on generating intermediary drawings (research question 1). We found no effects on retention of chemistry knowledge. This finding is not surprising because collaborative interactions are known to be less effective for the acquisition of simple knowledge than for complex knowledge (Kirschner, Paas, & Kirschner, 2010). Accordingly, we found higher learning gains on a test of chemistry knowledge transfer for students who received a model-focused collaboration script, compared to unscripted collaboration in a control condition. By contrast, students who received a drawing-focused collaboration script showed lower learning gains on the transfer test than the control condition.

In addition, because translating between representations is particularly difficult for students with low spatial skills, we investigated whether students’ spatial skills moderate the effects of model-focused and drawing-focused collaboration scripts (research question 2). We found that the advantage of the model-focused collaboration script and the disadvantage of the drawing-focused script were particularly pronounced for students with low spatial skills.

Why might the drawing-focused collaboration script have resulted in lower learning gains than unscripted collaboration? We propose three potential reasons. First, drawing intermediary representations might result in cognitive overload, particularly for students with low spatial skills. The purpose of the lab session in which we conducted our experiment was for students to learn to translate ball-and-stick models into wedge-dash structures. Hence, students were not yet proficient at generating drawings to translate between these representations. When asking students to draw on paper to plan their wedge-dash structures, students receive no support for doing so. Drawing without support can increase cognitive load (Schwamborn et al., 2011), especially for students...
with low spatial skills (Hegarty & Waller, 2005; Stieff, 2007), which in turn may impede learning. Second, focusing students’ collaborative interactions on drawings may reduce translation and sense making between the target representations. In our lab session, the goal was to learn to translate physical ball-and-stick models into wedge-dash structures. Because students in the draw condition were asked to plan their wedge-dash structures on paper first, they may have focused on copying the wedge-dash structure from the paper rather than reflecting on how it translates from the ball-and-stick model. Third, focusing students on generating drawings may reduce collaboration altogether. Even though the collaboration script prompted students to discuss intermediary drawings with their partner, students drew individually on their own piece of paper, which implies ownership. Focusing students’ collaborative interactions on representations that they “own” may discourage collaboration—for instance, a partner may be less inclined to modify someone else’s drawing. This might, in turn, reduce reflection on how the representations show concepts (van Dijk et al., 2014).

Why was the model-focused collaboration script effective, in particular for students with low spatial skills? We consider two potential reasons. First, focusing students’ attention on the ball-and-stick models may have increased collaboration. Students used a shared modeling kit to build the physical ball-and-stick models. Because students built models together, they “co-owned” the models, which may have encouraged them to modify the models while answering conceptual questions and constructing wedge-dash structures. Second, students with low spatial skills may particularly benefit from collaborative interactions that focus on physical models because their partner can help them spatially align the representations. To translate physical ball-and-stick models into wedge-dash structures, students have to mentally project the 3D model into a 2D plane for the wedge-dash structure. By definition, students with low spatial skills have more difficulties than students with high spatial skills in mentally rotating objects. Students in the model condition were prompted to collaborate with their partner to spatially align the physical ball-and-stick models with wedge-dash structures. The partners of students with low spatial skills may have helped them externally rotate the physical model to facilitate projection from 3D into 2D.

Limitations
Our results should be interpreted in light of the following limitations. First, causal inferences from quasi-experiments are generally limited because non-random differences between conditions may exist. Although we took all possible steps to ensure the equivalency of the conditions, an experiment with random assignment of individuals to conditions should replicate our results. Second, our experiment was situated in the context of a specific chemistry course. Even though we consider the naturalistic context of our experiment a strength in terms of external validity, future research should test if our results generalize to other STEM courses. Finally, our experiment focused on two specific representations. Therefore, future research should replicate our findings in other STEM domains that use different physical and virtual visual representations (e.g., 3D and 2D protein models in biology, geological layers and block diagrams).

Conclusion
Our results contribute to theory about collaborative learning. We provide novel insights into the mechanisms that may account for the effectiveness of collaborative activities on students’ learning with multiple visual representations. We isolated two possible mechanisms that focus students’ collaborative interactions on physical models or on generating intermediary drawings. Our findings suggest that, especially for students with low spatial skills, focusing shared attention on physical models is a mechanism through which collaboration may enhance students’ learning, whereas focusing on intermediary drawings is a mechanism through which collaboration might hinder students’ learning.

Further, our results make practical recommendations for the design of collaboration scripts and collaborative practices in STEM instruction. We recommend that collaboration scripts focus students’ collaborative interactions on shared models rather than on intermediary drawings. Further, we caution instructors against encouraging students to generate intermediary drawings when translating between representations, especially if they have low spatial skills. Our study shows that even a simple tweak of how students are prompted to collaboratively use shared representations can render a collaboration script more or less effective than unscripted collaboration.

References


Acknowledgments
This research was supported by the UW-Madison Graduate School, National Science Foundation through Award IIS #1651781, and Institute of Education Sciences, U.S. Department of Education, through Award #R305B150003 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the National Science Foundation and U.S. Department of Education. We thank instructors Judith Burstyn, Etienne Garand, Ive Herman, Judy Hines, and the teaching assistants who helped us implement the experiment. We also thank Hannah Bowman, Maggie Cram, Jainaba Corr, and Haengjung Ye for their conceptual and technical support.
Comparing First- and Third-Person Perspectives in Early Elementary Learning of Honeybee Systems

Kylie Peppler, Naomi Thompson, Joshua Danish, and Armin Moczek,
kpeppler@indiana.edu, naothomp@indiana.edu, jdanish@indiana.edu, armin@indiana.edu
Indiana University

Seth Corrigan, Collective Shift/LRNG, seth@lrng.org

Abstract: While prior literature has demonstrated that even young children can learn about complex systems using participatory simulations, this study disentangles the impacts of third-person perspectives (offered by traditional simulations) and first-person perspectives (offered by participatory simulations) on children’s development of systems thinking and biology learning. Through the lens of honeybee nectar collection, we worked with three first-grade classrooms assigned to one of three conditions -- instruction through use of a first-person perspective only, third-person perspective only, and integrated instruction -- to engage ideas of complex systems thinking. In each condition, systems concepts were targeted through instruction and assessment. The combined and third-person classrooms demonstrated significant gains while the first-person classroom showed gains that were not statistically significant, suggesting that third-person perspectives play a critical role in how children learn systems thinking. This work also puts forth a novel assessment design for young children using multiple-choice questions.

Introduction

From food webs to traffic to the respiratory system, complex systems are present in every facet of daily life. It is important to be able to reason and make decisions based on systems phenomena. However, we know systems thinking to be a difficult concept for people at all stages to grasp (Hmelo-Silver & Azevedo, 2006; Resnick, 1999). From an educational standpoint, the importance of these systems is two-fold. First, complex systems are all around us, prevalent in the natural and social world. Second, systems dynamics are quite generative and apply across content settings and domains despite being challenging for students to initially grasp, they are also a set of dynamics that are challenging for students to learn while also being generative and requisite for a deep understanding of most any domain of science (see Hmelo-Silver & Azevedo, 2006; Resnick, 1999). Given the value of systems understanding, researchers have been working to develop new insights into how we can make them more approachable. In particular, a number of researchers have proposed that we introduce students to systems concepts early in their academic careers, potentially transforming lifelong learning trajectories (Thompson, Peppler, & Danish, 2017; Danish, 2014; Assaraf & Orion, 2009; Grotzer & Basca, 2003).

One approach to engaging learners with complex systems concepts has been through computer simulations, where children can interact with visual representations of how systems elements interact (e.g., Danish, 2014; Hmelo-Silver, Eberbach & Jordan, 2014; Grotzer et al., 2015). A second successful approach has involved participatory simulations through which children physically act as agents within a system (Danish, Peppler, Phelps, & Washington, 2011; Colella, 2000). While both approaches have led to demonstrable learning gains for young children, few studies have investigated the unique contribution of each approach to learning. Is one generally more beneficial to young learners? Does each contribute to slightly different learning outcomes or are these approaches interchangeable?

Previously, we demonstrated that when combined within a curriculum, these approaches do provide students with unique opportunities to engage in discussions about the core systems concepts being studied (Danish, Thoroughgood, Thompson & Peppler, 2017; Thompson, Peppler, & Danish, 2017). The present analysis builds on this prior work with a similar aim of exploring how the third-person perspective offered by traditional simulations and the first-person perspective offered by participatory simulations work in different ways to support children’s systems thinking learning. The current project moves beyond our prior work by shifting from one integrated classroom to working with three first grade classrooms in a typical Midwestern elementary school to compare how students learned about systems thinking through the lens of honeybees across three conditions: first-person perspective only, third-person perspective only, and integrated (first- and third-person perspectives combined).

Honeybees as a complex system accessible to young children
The term “complex systems” is used to describe collections of inter-dependent and inter-related elements where the collection, or system, has properties that emerge from both the individual elements and their relationship to each other (Jacobson & Wilensky, 2006). In the case of honeybees collecting nectar, the initial topic of the current project, we can view the honeybees within a hive, the hive itself, and the flowers that the bees visit to collect nectar as a system. Honeybees collect nectar from these flowers, converting it into honey within the hive. As scout bees discover good sources of nectar, they return to the hive where they will perform a “bee dance” which indicates the direction and distance to the source of nectar, as well as the quality of the nectar source (Seeley, 1995). Other bees observe this dance and then set out in search of the identified flowers. If they in turn find an abundant supply of nectar, they return and repeat the dance. The result is not only an incredibly efficient nectar collection operation, but also a highly adaptive one with honeybees ceasing to visit flowers that are no-longer effective sources of nectar, and shifting rapidly to new supplies.

What makes this a complex system is not just the fact that it consists of different elements (i.e., the bees, the physical hive where they store their nectar, the flowers which may or may not have nectar), but also the inter-relatedness of these elements, and the different levels of analysis at which it operates (Hmelo-Silver & Azevedo, 2006). The different levels of the system are also the first place where we see a clear distinction between experts and novices. Novices tend to view a system such as this in terms of its superficial structures (e.g., the honeybee body parts) and behaviors (e.g., bees do a dance) instead of the functions of these behaviors and structures (e.g., the dance leads to faster nectar collection) (Hmelo-Silver, Marathe, & Liu, 2007; Hmelo-Silver & Pfeffer, 2004). One reason why the functions are so elusive may be that these functions often require an examination of the emergent properties of the system as a whole, rather than the local behaviors. One goal of the current project was, therefore, to help move young students from superficial descriptions of the system of honeybees collecting nectar to a more nuanced understanding of the functions that these different behaviors serve for the hive as a whole.

One successful approach to examining complex systems in education has been to focus on the process through which properties of the system “emerge” from the behaviors or properties of the individual elements or agents (Jacobson & Wilensky, 2006; Wilensky & Resnick, 1999; Wilensky & Stroup, 2000). In the case of honeybees collecting nectar, for example, we want to help children understand the way the hive as a whole is efficient at collecting nectar despite the fact that individual bees engage in behaviors that don’t immediately appear to be effective, such as spending time dancing (instead of collecting more nectar) or continuing to look for new flowers when the hive has already found some that represent a good source of nectar. The concept of emergence also pushes against the misconception that the queen bee directs all activity in the hive. Prior research has consistently shown that this kind of emergent property is even challenging for adolescents and adults to understand, particularly because it requires the ability to shift one’s perspective back and forth between “levels” of analysis, represented here by the individual bees and the hive as a whole (Hmelo-Silver & Azevedo, 2006; Wilensky & Resnick, 1999).

First- and third-person perspectives on early childhood systems thinking

While previous work has shown that although systems thinking is an advanced and complex topic, young children can begin to learn these concepts through play and embodiment (e.g., Danish, 2014; Assaraf & Orion, 2009). In particular, the concepts of feedback, a process that leads to increase or decrease in an action or behavior (Sweeney, 2012); iteration, a repetition and building up of an action or behavior (Wilensky & Resnick, 1999); and constraints, properties that make work within a system more difficult (Wilensky & Reisman, 2006) are crucial systems thinking concepts with which learners may struggle. As such, we built our assessment to address these three concepts. Drawing on activity theory (Engeström, 1987), we know that examining how joint activities are mediated is useful in understanding how learning occurs in these contexts and can provide insights into how the organization of activity might support students in engaging with different aspects of a system (Danish, 2014). For example, first person perspectives have proven powerful in helping students to engage with the constraints within a system, whereas a third person perspective has helped them to see the emergent properties more clearly (Danish, 2014; Danish et al., 2017). Capitalizing on these prior findings, the key design differences between the conditions in our current work is the perspective through which the activities are mediated: First- and third-person perspectives on the system. Table 1 summarizes our expectations for how the different perspectives would support students’ engagement with different aspects of the system.

The first-person perspective allows learners to take the role of an actor (i.e., a honeybee) in a system (i.e., collecting nectar from flowers and returning it to the hive). With this view, they can experience first-hand individual constraints that might arise for the actors, and see how other factors such as feedback may cause behavior adaptations in response to those constraints. However, patterns on a larger scale may be more difficult to track from this perspective.
We hypothesized that a first-person perspective of nectar collection could better illuminate the complex communication patterns that happen inside the hive and could be particularly important for our target age group, given the need to learn about complex systems from multiple analytic levels at the same time in order to fully understand the relationship between levels (Hmelo-Silver & Azevedo, 2006). While most prior work on participatory simulations has targeted older children, teens, and adults, this body of research fails to take into account the alignment between participatory simulations and the play activities of young children, who are already apt to explore topics of interest through play-acting and games (c.f., Youngquist & Pataray-Ching, 2004). In addition, it has been suggested that an agent-based perspective where students reason about the behaviors of individual agents within the system increases the potential of students to transfer their understanding to other systems (Goldstone & Wilensky, 2008).

Table 1: Hypothesized strengths of each model/perspective in the context of honeybees

<table>
<thead>
<tr>
<th>Targeted concepts</th>
<th>First-person views increase awareness of:</th>
<th>Third-person views increase awareness of:</th>
<th>Integrated first- and third-person views increase awareness of:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Honeybee behavior</strong></td>
<td>Reasons behind individual behaviors (e.g., why bees waggle dance)</td>
<td>How the hive as a whole adapts and pursues a single nectar source.</td>
<td>The impact behaviors that overcome challenges in the 1st-person perspective have on patterns in the 3rd person.</td>
</tr>
<tr>
<td><strong>Feedback loops</strong></td>
<td>The benefit one's own activities gain from feedback.</td>
<td>How feedback from a few bees can impact many more bees.</td>
<td>How providing feedback helps the hive as a whole, just as receiving feedback helps an individual bee.</td>
</tr>
<tr>
<td><strong>Iterative processes</strong></td>
<td>How repeated activities can look different at the individual level, depending on the context (e.g., does a flower still have nectar).</td>
<td>Repetition as a path to better results for the hive as a whole, in contrast to a hive without the same feedback loops and repeated patterns of behavior.</td>
<td><em>(No expected difference)</em></td>
</tr>
<tr>
<td><strong>Constraints</strong></td>
<td>Challenges individual bees face (e.g., finding flowers with nectar is challenging).</td>
<td>Challenges the hive as a whole faces (e.g., gathering enough food to survive through winter).</td>
<td>How individual and group constraints are interrelated.</td>
</tr>
</tbody>
</table>

By contrast, the third-person perspective allows learners to observe the actions of many actors in a system (i.e., all the honeybees that are searching for nectar) from a bird’s eye view. Here, rather than experiencing individual constraints, learners can visualize the larger impacts of constraints on the whole system (e.g., falling nectar levels in winter). The build-up of iterative behaviors into emergent patterns is also more salient from this removed viewpoint.

We hypothesized that the third-person perspective would better provide learners with a view of large scale patterns. Much of the previous work on systems thinking with children has focused on third-person computer simulations, often designed by the learners themselves (c.f., Wilensky, 2006; Wilensky & Resnick, 1999). As with the first-person work, this often occurs with older children and adults as complex computer programming is often involved. Previous work with the third-person simulation discussed here has shown success in helping very young children begin engaging with complex systems through the lens of honeybee hives (Danish, 2014). This view from the outside allows learners to track actions and outcomes and make connections to smaller scale behaviors that can be difficult to notice from a first-person perspective. It is also ideal for discussing iteration as students watch actions repeat and build.

**Guiding theory of learning: Activity Theory**

Here, we designed for learning by drawing on Activity Theory (Engeström, 1987; Engestrom, 2008), a theoretical framework grounded in the work of Vygotsky (1978) that focuses on learning as a mediated, social
process where learners move toward a shared goal. Activity theory provides ways to attend to the rules, tools, communities, and divisions of labor that shape the learning process. These mediators shape and define learners’ interactions with each other, the technology, and the learning goal. Using this theoretical perspective, we can focus on these intersections and the roles they play as children generate and transform their ideas about how systems work. From this perspective, it is understood that learning is socially mediated (see Vygotsky, 1987; Greeno, 2006). This informs the focus throughout the curriculum on group discussions and debrief. The contributions of individuals in a classroom impact the learning of the group as a whole, as well as individual students. In designing the curriculum, we used activity theory as a guiding theory for the design, paying particular attention to the system elements to build parity across the three conditions. Regardless of condition, the shared goals of each classroom were the same: to collect as much nectar as possible (either themselves or through a computer simulation), and to gain new understandings about honeybees along the way. The main difference across conditions was the perspective through which the activity was mediated.

General design-based research approach and research questions

Our general approach to the overarching work was a Design-Based Research paradigm (The Design-Based Research Collective, 2003) where we iterated on the designs of our intervention and assessments until we were able to conduct the present quasi-experimental study. The current study was designed to further refine our understanding of the differential contributions of first- and third-person perspectives on systems learning as well as the features of our designs that appear to support this learning. In order to accomplish this, we developed conjectures during our design process about what specific features of the tools we are developing we believe led to student exploration of or understanding of particular aspects of the content and then evaluate those conjectures as part of our summative evaluation (Sandoval, 2004). Throughout the design and implementation process, we worked regularly with children and teachers as co-designers, mutually determining the purpose, value, and interpretation of our software prototypes, physical materials, and curricular approach (Nelson, 2000). Application of this iterative process allowed the designs to progressively improve over the course of five iterations of technology and curriculum. To ensure that the proposed activity system effectively complements curricular and practical realities, we consult with practitioners to inform our design and development throughout the process.

We look to answer the questions: (1) How do children’s performances on a systems thinking assessment change before and after engaging in the BeeSim curriculum? (2) How do these changes differ between experiencing the curriculum through a first-person lens, a third-person lens, or a combination of the two? (3) How might first- and third-person perspectives work together and separately to support systems thinking learning?

The BeeSim designs

All three conditions in this study took place over about 12 days, or in about 8 to 9 hours of activity. In each classroom, we began by interviewing a sample of the students for a qualitative assessment of learning; this interview was repeated at the end of the implementation. Each implementation also began and ended with a 20-question multiple-choice assessment administered through a clicker system.

We wrote three versions of the BeeSim curriculum corresponding with the three conditions. The participating classroom teachers were involved in the curriculum design process and provided useful insights into classroom appropriateness. Each version of the curriculum followed the same progression from simple to more complex topics, introducing more nuanced accounts of systems thinking concepts -- feedback, constraints, and iteration -- along the way. Each condition began by introducing the basics of the honeybee body, then moved to the basics of bee nectar collection in a single hive, introduced competition by comparing two hives, introduced the waggle dance as the feedback mechanism, explored how feedback and system constraints were related, and considered the role of negative feedback (or the lack thereof). The typical pattern for a session began with teacher-led recaps and soliciting student predictions, was followed by the planned simulation play, and ended with a teacher-led debrief.

In the first-person condition, the planned simulation activity involved students using electronic bee puppets to collect nectar from larger than life flowers placed around a classroom. Large swaths of yellow fabric hung from the ceiling, creating “hive” spaces where students waited for their turn in the field. RFID tags in the flowers could be read by sensors in the bee puppets’ heads, allowing information to be transmitted between the bee and a central server. Lights on the bee’s body indicated how much nectar it was holding, how much energy it had, and the quality of a flower when checked. The curriculum progressed as indicated above, all framed through the first-person perspective. While students were occasionally asked about what an occurrence might mean for “the whole hive” or to think about more than just their individual bees, no activities or direct
information was given from a third-person perspective in order to separate how first-person learning activities impact learning about both first- and third-person perspectives.

In the third-person condition, the planned simulation activity involved students engaging with a computer simulation that represented idealized behaviors of honeybee hives. The interface could show either one or two hives. To compare patterns and the effects of the waggle dance, one hive could be set to “not dance” while the other was set “with dancing.” With all other variables held constant, students could observe and make predictions about the hives’ differences. Other adjustable variables included flower number, position, and quality. The curriculum progressed as indicated above, all framed through the third-person perspective. While students were occasionally asked about what an occurrence might mean for an individual bee, no activities or direct information was given from a first-person perspective in order to separate how third-person learning activities impact learning about both third and first-person perspectives.

The integrated condition incorporated parts of both the first- and third-person conditions. Students both played with electronic bee puppets and engaged with a computer simulation. We identified the form of simulation -- 1st person or 3rd person -- that we felt, based on prior experience, was most likely to engage students productively with the concept. For example, the first-person bee puppets were chosen to introduce the waggle dance because it would allow students to directly experience the role of the dance, whereas our experience suggested that students were more likely to see aggregate patterns through the third person simulation. Children in this condition also had access to a separate “playback” technology that allowed the movements of the bee puppets between the flowers and the hives to be played back in real time for the students to reflect upon. This provided an experience that was truly a blend of first- and third-person perspectives, as the bees represented the children’s individual actions, but were treated as a bird’s eye view.

Participants and data sources
The project took place in a public school in a mid-sized, midwestern city. The school population is 86.6% white, with 41.8% receiving free or reduced lunches. The school, and our particular classrooms were evenly split in terms of gender. Data collection occurred in the schools three first grade classrooms; we worked closely with each teacher who led activities and approached the curriculum in ways appropriate for her students. We compared performance and change across two time points for the three conditions: first-person (n=19), third-person (n=21), and integrated (n=20) perspectives. To compare the differences between the three conditions, we collected multiple forms of observation and assessment data. All sessions were video and audio recorded. All students took two pre- and post-assessments before and after the 12 day intervention. The first asked students to draw how bees get food. The second was a 20-question multiple-choice assessment given through a clicker system. Over three design and testing iterations, we created this assessment to ask eight “simple” biology-based questions (e.g., This forager bee just came out of the beehive. Its job is to collect nectar. What will it do next?) and twelve “complex” systems-based questions (e.g., This bee saw a waggle dance that said this flower had a lot of great nectar, but all the nectar was gone when it got there! If other bees saw the same waggle dance, what would they do?). All questions and answer options were read out loud at least twice to alleviate concerns about varying reading levels. It is important to note that due to concerns of test fatigue with young students -- most first grade students do not often take multiple-choice assessments -- we were unable to tease apart some of the more nuanced concepts we hypothesized would differ across conditions. Also, a random sample of 20% from each condition was chosen to participate in pre- and post-interviews to provide qualitative evidence of understanding.

This paper focuses on the results from the pre- and post- multiple-choice questions. We used a latent variable modeling approach (Skrondal and Rabe-Hasketh, 2004), employing the unidimensional Rasch model to estimate impacts of the BeeSim instructional content. Two general rounds of analysis are presented. The first round of analysis groups all students regardless of condition, responding to RQ1. The second round of analysis – treatment specific analysis - disaggregates students into their respective condition groups in order to facilitate a three-way comparison of the mean student growth for each group across the pre-test and post-test, responding to RQ2. Use of the Rasch model employs students’ response vectors for the respective assessments in order to estimate student abilities measured in logits, and standard errors. Using students’ resulting parameter estimates at pre-test and post-test, and a difference-in-differences, we employed the simple t-test in order to determine the significance of change in performance from the pre- to post-assessments.

Results
Results from a total of 60 students were used in the analysis. A total of three students were dropped from the analysis because they were missing responses to more than half of the test items at pre-test and post-test. A summary of the following results can be found in Table 2 below.
When all three conditions were grouped together (N = 60), the mean student ability estimate at pre-test was -0.472 logits with a standard deviation of 0.657. The mean standard error for the pre-test ability estimates across the sample, is 0.526 logits. By comparison, the mean student ability at post-test was +0.188 logits with a standard deviation of 1.351 logits, and an average standard error of 0.606 logits. Thus, ability estimates at post-test were higher on average and more dispersed (exhibited a larger variability across the sample) than those at pre-test. Use of a paired sample t-test confirms that the difference in ability estimates is significant. The difference in mean ability estimates between pre-test and post-test administrations is more than half of a logit, +0.660 logits, with t = 3.897 (Confidence Interval: 0.321, 0.999) and p < 0.0001. Using Cohen’s d to calculate the associated effect size, this difference in means is associated with an effect size of 0.621 – appropriately categorized as a ‘medium’ sized effect. These results address RQ1 which asked about general gains across all conditions on the systems thinking pre- and post-measure.

To address RQ2, we also compared the performance of each condition. As a first step, students’ ability estimates across the three groups were compared to ensure they were equivalent. In each case the differences in the three groups’ mean ability estimates were not statistically significant (p > 0.05). This supports the claim that the three groups were equivalent with regard to their ability estimates at pre-test, though we note it does not guarantee that the groups were equivalent with regard to other baseline characteristics.

On average, students in the classroom receiving the first-person condition (n = 19) did not exhibit significantly different abilities between the pre-test and the post-test administrations of the BeeSim assessment. At pre-test, the mean ability estimate of the group was -0.291 logits with a standard deviation of 0.742 and a mean standard error of 0.533 logits. At post-test, the mean ability estimate was -0.103 logits with a standard deviation of 1.479 and a mean standard error of 0.631 logits. Application of the t-test with 18 degrees of freedom resulted in t = 0.650 and p < 0.523.

In the third-person condition (n = 21) students had significantly higher ability estimates at post-test than at pre-test. At pre-test, the mean ability estimate for this group was -0.457 logits with a standard deviation of 0.695 and a mean standard error of 0.152. The mean ability estimate at post-test was 0.470 logits with a standard deviation of 1.548 and a mean standard error of 0.338 logits. Use of the t-test resulted in t = 2.706 with 20 degrees of freedom, and p < 0.05. Applying the function for Cohen’s d, there was a medium estimated effect size found of 0.77.

The largest change in ability estimates across the three groups was exhibited by students in the integrated condition (n = 20). Among these students, the average ability estimate at pre-test was -0.658 logits. Those estimates have a standard deviation of 0.492 and a mean standard error of 0.110 logits. At post-test, the mean ability score for the group had risen to +0.169 logits with a standard deviation of 0.949 and a mean standard error of 0.212. The t-test results were statistically significant with t = 3.882 with 19 degrees of freedom and p < 0.001. The magnitude of the effect was large with Cohen’s d = 1.09.

Table 2: T-test results for change from pre- to post-assessment

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pre (Logits)</th>
<th>Post (Logits)</th>
<th>Change (Logits)</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>60</td>
<td>-0.472</td>
<td>0.188</td>
<td>0.66 (p &lt; .001)</td>
<td>0.621</td>
</tr>
<tr>
<td>First-Person</td>
<td>19</td>
<td>-0.291</td>
<td>-0.103</td>
<td>0.188 (p &lt; .5)</td>
<td>NA</td>
</tr>
<tr>
<td>Third-Person</td>
<td>21</td>
<td>-0.457</td>
<td>0.470</td>
<td>0.927 (p &lt; .05)</td>
<td>0.77</td>
</tr>
<tr>
<td>Integrated</td>
<td>20</td>
<td>-0.658</td>
<td>0.169</td>
<td>0.827 (p &lt; .01)</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Additionally, we categorized the multiple-choice items into “simple” and “complex.” Simple items were those pertaining to non-systems thinking content, such as honeybee biology (e.g., honeybees have a head, thorax and abdomen) or behaviors (e.g., honeybees take on different jobs and responsibilities throughout their lifecycle). Complex questions addressed the systems thinking concepts feedback, constraints, and iteration. We calculated the average percentage of correct answers for each category and compared the three conditions. The integrated condition had the largest gain on the simple items, moving from 43.55% correct to 63.29% correct.
The Simulation condition had the largest gain on the complex items, moving from 46.67% correct to 59.34% correct. See Table 3 for a summary of these results.

Table 3: Change in average percent correct from pre- to post-

<table>
<thead>
<tr>
<th></th>
<th>Simple Pre</th>
<th>Simple Post</th>
<th>Gain</th>
<th>Complex Pre</th>
<th>Complex Post</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>First- Person</td>
<td>48.61%</td>
<td>49.47%</td>
<td>0.86%</td>
<td>42.59%</td>
<td>49.62%</td>
<td>7.03%</td>
</tr>
<tr>
<td>Third- Person</td>
<td>46.67%</td>
<td>59.34%</td>
<td>12.68%</td>
<td>36.90%</td>
<td>55.20%</td>
<td>18.29%</td>
</tr>
<tr>
<td>Integrated</td>
<td>43.55%</td>
<td>63.29%</td>
<td>19.74%</td>
<td>30.76%</td>
<td>46.29%</td>
<td>15.53%</td>
</tr>
</tbody>
</table>

**Conclusion and discussion**

These findings suggest the third-person and integrated conditions provided better systems thinking learning outcomes, although all three conditions made gains from the pre- to post-assessment. Students in the third-person condition also performed best on the complex items, although the gain was only slightly more than the integrated condition. While this suggests the third-person perspective might be the most important, our ongoing qualitative analyses are beginning to indicate that the first-person perspective supports some important conversations and thus learning gains that are not yet measured through this multiple-choice assessment that may continue to complexify these early findings.

In addition to these results on the multiple-choice assessment, we hypothesize that moving back and forth between first- and third-person perspectives, as in the integrated condition, provides some additional insight and experiences not reflected in the multiple-choice scores. In our presentation, we will also share further analyses of the other assessment items as well as the classroom video data to reveal the nuances provided by the three conditions.

**Scholarly significance**

This work suggests three main outcomes and implications for systems thinking education research. First, early elementary students are capable of learning systems thinking concepts, particularly through third-person perspectives. This is particularly compelling given the costs and easy scalability of digital simulations such as these. It may seem intuitive that children in this age range would have difficulty reasoning about systems from an outside perspective, but our work demonstrates that this viewpoint was especially beneficial for learning outcomes. This is important for early childhood educators in particular, as freely-available curriculum and tools developed for the third-person perspective could help educators begin to introduce systems thinking earlier and more often, through embodied, playful techniques and familiar, high-quality biology content.

Second, there seem to be additional benefits to engaging in a system through multiple perspectives. Students in the integrated condition did significantly better on the post-assessment than the pre- with an effect size larger than the third-person condition. This suggests that further research and design iterations may more fully utilize the affordances of moving back and forth between first- and third-person perspectives. That said, the current research seems to locate the driving factor for these gains, particularly in learning about complexity, are fueled by third person perspectives. As a result, future iterations of the BeeSim platform is working to investigate real-time data visualizations of first-person player activity through our work in a new system, called AntSim, through a uniquely developed indoor positioning system. This work is the focal point of our future research and publications. One hypothesis that we have around why third person perspectives may be driving these learning outcomes is that is to quicker to watch multiple rounds of bees foraging for nectar and to see the emerging behaviors via a computer simulations (and can even be played in a fast forward fashion) than to play the participatory simulation bee puppet game with a group and develop the sufficient expertise in the game so that the emergent patterns can be seen in the same way. Consequently, the third-person perspective condition may have had more time to debrief more fully and deeply engage the content across more cycles of play than in-person groups. Future studies may wish to design comparisons between first and third person perspectives with this limitation of the group comparisons in mind. Integrated maybe didn’t get to dive as deep into simulation as would have been necessary for better results.
Lastly, further research should continue to tease apart complex and simple concepts in systems thinking and additional types of systems thinking content as it may be that emergence is particularly well suited for third person perspectives while other systems thinking concepts may be more well-suited for first-person perspectives. The differences in performance between the conditions on these simple and complex measures suggests that first- and third-person perspectives may impact levels of systems thinking differently. Defining these differing impacts more fully may have implications for the design and progression of systems thinking curriculum in the future.

References

Acknowledgments
This material is based upon work supported by the National Science Foundation under Grant No. 1324047 awarded to Kylie Peppler, Joshua Danish, and Armin Moczek. 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 National Science Foundation. Thank you to Janis Watson and the many teachers and students who made this work possible.
Mobilizing Learning Progressions for Teacher Use: Examining the Utility of Outside Learning Progressions in Task Co-Design

Erin Marie Furtak, University of Colorado Boulder, erin.furtak@colorado.edu
Kelsey Tayne, University of Colorado Boulder, kelsey.tayne@colorado.edu

Abstract: While many learning progressions have been developed in science education, few studies have examined the utility of these progressions as tools for teachers in contexts outside of their original design. This paper is a case study examining how one group of high school biology teachers drew upon two different learning progressions (evolution and carbon cycle) over the course of one academic year as they designed and enacted formative assessment tasks, and then interpreted responses to the tasks. We find that the progressions were taken up differently and may have provided central ideas for the curricular units but primarily acted as tools for facilitators to frame discussion around student ideas. Our findings suggest the progressions appeared not to be ready-to-use tools for teachers in these new contexts.

Introduction

Learning progressions – representations of the sequential development of student ideas and scientific practices within core content domains (Corcoran, Mosher, & Rogat, 2009) – have been the focus of much science education research in recent years (e.g. Duschl, Maeng, & Sezen, 2011). Scores of learning progressions have been developed across multiple disciplinary core ideas and scientific practices in science, and these progressions have been developed and used for a number of different purposes, including curriculum design, assessment development, and modeling of student growth.

Among the lines of research that have developed around learning progressions is a set of studies that have explored, in a qualitative sense, how learning progressions can serve as tools to support teacher design of formative assessments (Briggs & Peck, 2015), their ability to diagnose student thinking (Furtak et al., 2016), and the ways in which they interpret student response data linked to the progressions (Alonzo & Elby, 2014). These results, while encouraging, have each focused on teachers using learning progressions in contexts closer to that in which they were originally designed, either working with similar curriculum materials (e.g. Furtak & Heredia, 2014) or facilitated by designers of the progression (Alonzo & Elby, 2014). As a result, the field has few images from research as to how the multitude of published learning progressions might support teachers outside of the context of the progressions’ original design and development. This gap in our collective understanding represents a significant limitation in understanding the scalability of this line of research. This paper presents a case study of a group of high school biology teachers working with two learning progressions to support their engagement in formative assessment task design, enactment, and interpretation.

Background

We frame our research from a situative perspective (Greeno, 2006), taking learning as changes in participation in practice, and teacher and student participation in classroom activity as coordinated by sets of tools and resources that embody particular design principles intended to structure participation in particular ways (Akkerman & Bakker, 2011). In the following sections, we provide relevant background to learning progressions, arguments for using them directly with teachers, and the ways in which learning progressions may support formative assessment task design and enactment.

Learning progressions

Learning progressions are hypotheses about the ways in which student thinking develops over a period of time (NRC, 2007). Several researchers have identified types of learning progressions, including Wilson (2009), Duschl, Maeng and Sezen (2011), and Lehrer and Schauble (2015). The variations in designs range from those with a small (e.g. within one unit) to a large (e.g. across a K-12 span of time) grain size; those that include students’ prior experiences and what some might call ‘misconceptions’ and those that focus on the ways in which students learn correct ideas in developmental progression; and those that focus on a single dimension of student understanding versus those that are multidimensional (Catley, Lehrer & Reiser, 2005).

Regardless of the design features of a learning progression, we emphasize that these progressions should not be viewed as developmentally inevitable (Alonzo & Gotwals, 2012), but rather lay out how student thinking and engagement in practice might unfold in the presence of a particular set of learning experiences,
such as a curriculum that takes students through particular stages of understanding. From this perspective, learning progressions, while likely helpful tools in the development of particular curriculum materials and assessments, likely will not maintain the same meaning when traveling from one locus to another, but will take up different meanings in new contexts (Star, 2010). In fact, teachers might not perceive learning progressions developed outside their local context as useful (Furtak, 2012).

From this framing, a central question about learning progressions as hypotheses about student understanding and engagement in practice is not whether or not learning progressions are ‘right’ in terms of capturing student learning in all contexts, but the extent to which they serve a particular purpose in a given context of use (Lehrer & Schauble, 2015). In the case of our own research, we ask questions as to whether learning progressions might serve as useful resources for teachers when taken away from the original context of their design. Given the amount of resources dedicated to these progressions and the materials they support, it is important to address the emergent question of whether progressions offer support when used in new settings, and if so, what kind of activity these resources support. We describe one such possible setting - a teacher learning community focused on formative assessment design - in the next section.

Formative assessment design cycle

We have engaged in a series of projects in partnership with teachers, schools, and districts to examine the ways in which communities of teachers draw upon the information included in learning progressions to inform processes of formative assessment task design and enactment. We define formative assessment as the tasks and tools that teachers use to elicit student thinking and organize classroom participation structures around attending and responding to student thinking, as well as the processes in which teachers and students engage to make learning goals explicit, share thinking, and provide feedback to move learners forward (Bennett, 2011).

Our program of research follows assertions made by many in mathematics and science education who argued that learning progressions, while important tools for researchers, might also support teachers’ classroom practice. For example, Heritage et al. (2009) suggested that learning progressions might help to concretize the ‘next steps’ part of formative feedback that can be so elusive to teachers, even once they have diagnosed student thinking. Similarly, Bennett (2011) argued that a learning progression could help teachers distinguish among the ideas students express as they learn. Furtak (2012) suggested learning progressions might serve as frameworks for teachers to design formative assessments help them interpret and respond to student ideas during instruction.

Drawing on these and other studies (e.g. Borko et al., 2008), we developed the Formative Assessment Design Cycle (FADC) as a multi-step process that supports teachers in the development of formative assessment tasks. The cycle begins with facilitators walking teachers through a learning progression to Explore Student Thinking (1), using a learning progression to categorize and interpret student work samples. Next, teachers identify ideas on a learning progression that they would like to assess during their instructional units, and Design Tasks (2) to specifically elicit those ideas, anticipate the ways students might respond to the task, and rehearse the types of feedback they would provide to different types of ideas (Horn, 2010). Learning progressions can provide fundamental support in this step, as they provide a continuum of ideas students may hold as they progress through a sequence of learning. The fourth step has the teachers Enact Tasks (3) in their own classrooms, using a learning progression either implicitly or explicitly to interpret student thinking on-the-fly. Finally, teachers come back together to Reflect (4) on classroom enactment by looking at student work together and using a learning progression to categorize groups of student responses and plan next steps to move students forward on the learning progression. This feedback is discussed in multiple timeframes (Wiliam, 2007), such that teachers identify not only what they will do in the next class session, but also how they will draw upon this information to support students for the rest of the unit, and academic year. At the same time, teachers reflect on the task itself, identifying the extent to which it elicited student ideas on the learning progression, and how it might be improved and revised for the next year.

In this paper, we used previously-developed learning progressions to guide formative assessment task design, enactment, and reflection in the FADC, and collect multiple sources of evidence to respond to the following research question: How did a biology teacher learning community use two learning progressions during enactment of the formative assessment design cycle? How can learning progressions originally designed for other contexts support teachers in the design and interpretation of classroom assessments?

Method

This paper draws on data collected as part of a larger study that explores the way that high school science teachers design, enact, and reflect upon formative assessments with the support of multiple learning progressions. We worked with previously-published learning progressions to support teachers’ formative assessment design at multiple high schools across content areas. In the present analysis, we focus on the case of
Learning progressions
The first learning progression, originally developed as part of the Carbon Time curriculum, traced energy and matter cycling in socio-ecological systems from K-12 and beyond (Figure 1; Mohan et al., 2009). Lower-anchor ideas on the learning progression begin with informal accounts that students often have when they enter school, based upon their observations of plants and animals, decomposition, and flames consuming fuel. Levels of the progression move from more macroscopic observations, plant and animal growth, toward a more microscopic account of the molecular processes of photosynthesis, respiration, and combustion. Ultimately, at the top level, students are able to examine the ways in which carbon is involved in these processes at the molecular level.

<table>
<thead>
<tr>
<th>Upper Anchor</th>
<th>Carbon-transforming process</th>
<th>Generating organic carbon</th>
<th>Transforming organic carbon</th>
<th>Oxidizing organic carbon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific accounts</td>
<td>Photosynthesis</td>
<td>Biosynthesis</td>
<td>Digestion</td>
<td>Biosynthesis</td>
</tr>
<tr>
<td>Macroscopic events</td>
<td>Plant growth</td>
<td>Animal growth</td>
<td>Breathing, exercise weight loss</td>
<td>Decay</td>
</tr>
<tr>
<td>Lower anchor: Informal accounts</td>
<td>Natural processes in plants and animals, enabled by food, water, sunlight, air, and/or other things</td>
<td>Natural process in dead things</td>
<td>Flame consuming (fuel)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Carbon Time learning progression (Mohan et al., 2009, p. 684).

We also worked with the Elevate learning progressions, originally designed for use in high school biology (Furtak & Heredia, 2014). Building Mayr’s framework for natural selection (Mayr, 1982), the Elevate progressions integrate ideas from ecology (e.g. biotic potential and population stability), genetics (variation, inheritance), and the consequences of those ideas on differential survival and reproduction of individuals within populations over long periods of time (Figure 2). The vertical levels represent how student thinking can develop over time, from lower-anchor ideas students bring to school toward upper-anchor, scientifically-accepted ideas.

Participants
The five participants in this study were all of the teachers responsible for one or more biology courses at Prospect High in the 2016-2017 academic year. These teachers ranged from 3-19 years of classroom experience, and all had college degrees in biology. We were also participants in the community as facilitators. The first author is a former high school biology teacher who has supported science teachers in formative assessment design and enactment for more than a decade. The second author has a background in K-12 informal science and environmental education and previously worked with teachers on climate change education.

Professional development approach
The authors, either together or separately, met with the biology teachers at their school twice each month during the 2016-2017 academic year to introduce the learning progressions, guide the process of formative assessment task design, and provide structure as teachers categorized student work samples and identified next instructional steps. The meetings were designed to follow the Formative Assessment Design Cycle as described above.
Sources of data
We used multiple sources of data to triangulate our findings for this paper, which reflect our experiences at 11 different on-site meetings at Prospect High school year. Our primary source of data is in-depth fieldnotes created by the two authors at each of these meetings. We also kept copies of teacher learning community meeting agendas, the two learning progressions, copies of teacher-designed formative assessments, and copies of student work. We also took pictures of teacher-created artifacts during on-site meetings. In addition, we conducted interviews with teachers in which we explicitly asked them about their impressions of the learning progressions and how they supported their diagnosis of student thinking.

Analytic approach
Working with fieldnotes and transcribed teacher interviews, the second author developed an initial, open coding approach in which she identified general themes and patterns in the data focused on the ways that teachers used the learning progressions to support their task design. Both authors then met together to discuss these themes and patterns, and then developed a research memo format that would inform the next pass through the data. Both authors then read and analyzed all fieldnotes, creating separate research memos that summarized the ways in which the learning progressions were used in the different phases of the FADC and, as an additional step to search for disconfirming evidence of instances in which the learning progressions could have been used, but were not. These memos were created at a low level of inference and were cross-referenced with the data to create an audit trail. The authors met and discussed the memos in detail, and the second author also pulled teacher responses to interview questions specifically about the learning progressions to discuss and triangulate with the data. During these discussions, the authors recorded all of their adjudicated claims, and kept notes on overarching themes that emerged from the data. We then used these memos to create drafts of our results section, following Erickson’s (1986) guidelines for particular description and general commentary.

Results
Teachers’ formative assessment task design work spanned 11 meetings across the academic year, with a focus on the matter and energy cycling progression in the fall semester, and the natural selection progressions in the spring. This timing was dictated by teachers’ usual progress through the SEPUP curriculum, the adopted curriculum resource at Prospect High. We present our findings according to the teachers’ focus on each learning progression individually, and then draw contrasts and comparisons across the progressions.

Matter and energy cycling: The ‘Carbon Time’ learning progression
The first learning progression we used with the teachers, the Carbon Time progression, was intended to support teachers’ formative assessment task design and enactment in their fall ecology unit. This process began in the Explore student ideas phase with teachers first developing what we called a ‘local scope and sequence’, which they created to represent the sequence of concepts and activities they taught with the support of the SEPUP curriculum. Although the university-based facilitators introduced the learning progression to the teachers in an early meeting, the teachers did not use it directly to adapt or revise this scope and sequence document. Mostly, the facilitators used the learning progression to support a discussion of student ideas as a range or spectrum, not just right or wrong answers. At Meeting 2, a facilitator described the learning progression to the group as a way of understanding where students’ ideas are and supporting students in developing new ideas. At Meeting 3, a facilitator referred to the Carbon Time learning progression as having significant “upward freedom” for assessment purposes because of the wide range of ideas that the learning progression presents.

As the semester unfolded, the facilitators moved teachers into the design formative assessment tasks phase, and teachers created two activities to use in their classrooms, both adaptations from the resources designed as part of the Carbon Time curriculum. Facilitators had reviewed items from the assessment linked to the Carbon Time curriculum prior to the meeting, and selected possible tasks to share with the teachers. One of these tasks consisted of questions about energy and matter in a burning match, and the other included questions about matter and energy cycling in a tree. Despite these materials’ link to the learning progression, there was no explicit reference to the progression itself when teachers were adapting the tasks for use in their classrooms in our meetings together. For example, in the iterative design of the Energy and Matter in an Oak Tree formative assessment, teachers discussed the value of having multiple questions that addressed a similar scientific idea so that teachers can better understand what students are thinking. In the final version, shown in Figure 3, students were given several inputs (air, sunlight, water, soil) and asked where the mass is coming from as an oak tree grows from an acorn and where energy is coming from. Students were asked to circle if most, some or none of the mass/energy came from those inputs.
Figure 3. A task for matter and energy cycling, teacher-designed “Oak Tree” task on the left and a formative assessment task for natural selection unit “Peppered Moth” task (Furtak, 2012) on the right.

Teachers then enacted these tasks with students and brought student work samples back to a meeting to Reflect and Identify Next Steps. Margaret brought a copy of the Carbon Time progression to help her sort student work, noting that she had already tried to use it to make piles of student work but wasn’t able to align the categories in the progression with the student responses to the Burning Match task. We encouraged her to work with her colleagues to develop their own system for making piles of student work that made sense to them. While one pair of teachers made just two piles – those that were right and wrong – the other groups made more descriptive categories, including “matter consumed,” “matter transformed/converted,” and “into air/smoke.” They then discussed instructional supports that might be provided for the students in each category, drawing on the categories they had created. Later, when teachers interpreted student response patterns to the Energy and Matter in an Oak Tree task, teachers and facilitators created piles of student work again but did not reference or attempt to use the Carbon Time learning progression to do so. In discussing formative assessment design and enactment in one particular meeting, a facilitator referenced the learning progression saying that a written artifact can be compared back to the learning progression. Despite the challenge to use the learning progression to sort student ideas, teachers indicated that creating the piles and sorting student data was valuable for their practice. One teacher shared that she did not feel that the sorting process was valuable at first but that afterwards it helped her to better understand specific student responses so she could support her students.

Natural selection: The ‘Elevate’ learning progression

We began Setting Learning Goals and Exploring Student Ideas for the Evolution unit in the spring semester by introducing the natural selection learning progression at Meeting 8, and then supporting teachers in adapting their own local scope and sequences for the natural selection unit at Meeting 9. The teachers had multiple separate scopes and sequences they created as part of this process which, they later reflected in interviews, were likely influenced by the Elevate progressions. That said, like the Ecology planning process, teachers did not directly use the Elevate progressions to create their scope and sequences, but rather formed these on the basis of their prior instructional approaches and curriculum materials. Teachers discussed student ideas about how organisms change in response to the environment, a common misconception in this domain (Shhtulman, 2006), but they did so prior to examining how these ideas were incorporated into the learning progression.

Teachers also explored student ideas about natural selection by analyzing student responses (collected from another school) to a formative assessment task developed as part of a previous study (Furtak, 2012), and watching a video of high school students discussing their responses to that task. Facilitators asked teachers to create a spectrum of student ideas from what they felt were more correct to less correct. In this conversation neither the facilitators nor the teachers readily referred back to the learning progression for this activity.

Next, as teachers and facilitators Designed and revised formative assessment tasks, they used a variety of resources including the Elevate progressions, state and district standards, activities they had used in the past, their textbook, and a concept map of the big ideas covered in the unit developed by one of the teachers. Teachers considered designing a sequence of formative assessment tasks around the Elevate progressions to help them decide where to put the tasks in the sequence of their unit, but ultimately did not end up creating any formative assessment tasks this way. After reviewing several possible tasks to use with the teachers in this phase, the facilitators selected a task from a prior study as it was relatively simple, they had student response
data for it to examine, and could also share video of teachers using it with students. Ultimately, they used the task facilitators had provided them without changing it – the Peppered Moth, shown in Figure 1 - and another task, Climate Change Extinction, developed after teachers looked at the results of the Moth Task.

Similar to the carbon cycle unit, when reflecting on enactment, teachers analyzed student data from the Peppered Moth formative assessment by creating multiple piles of student responses based on student response data, not the learning progressions. Piles were related to students expressing ideas about organisms being able to change themselves, or ideas about differential reproduction. These ideas, which are related to the Elevate progressions, were developed from the task itself, but by explicitly using the Elevate progressions.

Cross-case analysis: Relative utility of progressions in supporting task design
Teachers described the learning progressions as a “menu” or something to provide “benchmarks” for the unit as they designed formative assessment tasks. While all of the teachers mentioned the Elevate progressions, only one mentioned the Carbon Time progression, likely because we interviewed teachers at the end of the year when they had just used the Elevate progressions. Ron felt the Carbon Time progression was less clear; when asked directly about the Carbon Time progression, Jim noted that there just hadn’t been enough time to use it, saying “we weren’t able to experiment with the carbon time stuff just because the time constraints are nutty.”

Teachers explicitly pointed to the Elevate progressions as a source of big ideas. Erika reflected that, “We used the evolution one that you guys showed us, where it showed us how it would develop over time and I mean, really we just kind of made a list of what were the big ideas.” Sarah also described the Elevate progressions as supporting her in thinking about student ideas and how these ideas progress through a unit, stating, “I think it’s useful to know where the students should be and then where we can take them with that work, the learning progression like, what, how do they follow that and where should they be at the end so that we can go back and see what we need to know from them in the beginning and then base our curriculum on that, or what we’re going to teach on that.” However, teachers’ descriptions of the learning progression as a resource for considering big ideas was somewhat vague. For instance, when asked about what features a teacher found more and less useful, Jim responded, “Well, the evolution [progressions], I don’t have the specifics, I’d have to take that out of the Google Drive, but it was a good introductory start.”

Ron suggested the Elevate progressions had influenced his classroom practice, noting that “I spent more time on the struggle for survival this year versus other years, basically making that connection between overproduction of everything from insects to elephants and that builds in a struggle for survival.” In terms of how the Elevate progressions supported unit design and sequencing, Sarah said that it broke down “…what [students] would need to know first, like natural selection, and then they understand that and what they need to know about that to get to the end point.” Ron mentioned the challenge of not having specific lessons to connect to the learning progression, noting that he felt the progression needed “…to be put in context of a lesson that would be delivered and tied to it so that there is a clear lesson formative matching. If you have the formative but there’s no lesson attached to it, then you’re going to have to cobble together a lesson.”

Discussion
Taken together, our case analysis – while admittedly with one school and a small sample of teachers – suggests some areas in which the progressions were directly useful to teachers, and other areas in which they could have been used but were not. We first summarize these main conclusions, and then identify future areas for research.

Lack of alignment between progressions and local scope and sequence
The SEPUP curriculum in place at our partner high school was built on the concept of spiraling, where core ideas re-occur throughout a year of study, and are treated with increasing complexity as the school year progresses (Bruner, 1961). In this environment, teachers created scopes and sequences to guide their instruction that were more tied to the curriculum materials they were accustomed to using, rather than adapting them to the learning progressions. Since many learning progressions are developed with the intention of supporting a process of curriculum design, this finding may only be an issue when teachers are using resources developed with a learning progression to supplement other instructional resources created with other designs and progressions in mind. Ultimately, given the piecemeal approach that many high school teachers take when selecting materials from multiple sources to support their instruction of different core ideas, a lack of alignment – whether real or perceived - between teachers’ current curriculum materials and those associated with learning progressions may ultimately limit the ways in which progression-based resources are adopted in classrooms.

Suites of tools to support teacher practice
We note that in both units we studied, the teachers used one formative assessment task that was provided to them as part the learning progression resources, and then adapted materials to create a second formative assessment task. In each case, the tools were taken up into classroom use without direct reference to the learning progressions, and the progressions were also not used directly to interpret student responses. Given perspectives about how teachers need suites of related tools to support their implementation of new instructional practices (Thompson, Windschitl & Braaten, 2013), this suggests that while learning progressions themselves may not always be directly useful to teachers, providing teachers with resources linked to the progressions might help to support them in instructional approaches related to the ideas the progressions contain. When teachers were sorting student responses to the Moth Task, for example, teachers created piles based on the student responses themselves, but these piles were directly related to the learning progression given the close link between the structure of the Moth task and two of the Elevate progressions.

Learning progressions as meditational tools for facilitators

Although the learning progressions themselves were not directly used in the ways we might have expected, we acknowledge that they did inform the ways in which facilitators described student thinking as a spectrum, pushing beyond binary framing of student ideas. Furthermore, the progressions guided the facilitators in selecting instructional resources to use to inform formative assessment task design (the Burn Match and Peppered Moth). In this sense, the tools mediated the work of the facilitators and the teachers, suggesting an additional use for these resources: as tools to guide the design of professional learning experiences for teachers.

Grain size

There were a number of key differences between the learning progressions, namely that the Carbon Time progression spanned multiple years of student learning, whereas the Elevate progressions represented student thinking as it might unfold within a unit of instruction in one year of a high school course. Teachers found the Elevate progressions more directly useful in informing their planning; however, this might be because the representation of the Elevate progressions was more closely related to the types of scope and sequences they used to plan units, rather than having such a zoomed-out view on student learning. Our study, while small and exploratory, suggests that the progressions that now underlie the NGSS may also be difficult to translate into direct action as teachers design and interpret tasks. Such progressions will need additional tools and resources such as sample tasks, rubrics, or other interpretive frameworks to be useful at the classroom level.

Conclusion and future directions for research

Ultimately, our study generates several questions about future directions in bringing learning progressions directly to teachers. First, it points to the importance of considering learning progressions as belonging to sets of resources, all of which might support teacher learning and task design. It also tempers our expectations around the extent to which learning progressions in and of themselves are directly useful for teachers, and suggests that future researchers may consider how they can inform the design of teacher activities in professional learning experiences. From this perspective, although learning progressions might continue to be developed, we may come to reposition them in the field – rather than an end in and of themselves – as a representation of cognition that then serves a foundation for the design of subsequent resources (e.g. Pellegrino, Chudowsky & Glaser, 2001), and then these sets of resources could travel to new contexts, rather than the progressions in and of themselves. Finally, in engaging in learning progressions research, our findings suggest that researchers need to consider the contexts, implications, and potential uses of the progressions they create in order to strive to produce research that has meaningful implications for use in guiding science teaching, learning, and assessment.

References


**Acknowledgements**

This material is based upon work supported by the National Science Foundation under Grant No. 1561751. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Designing From Outer Space: Tensions in the Development of a Task to Assess a Crosscutting Concept

Erin Marie Furtak, University of Colorado Boulder, erin.furtak@colorado.edu
Torsten Binder, University of Duisburg-Essen, torsten.binder@uni-due.de
Kate Henson, University of Colorado Boulder, kate.henson@colorado.edu

Abstract: New visions of science learning that integrate disciplinary core ideas, scientific practices, and crosscutting concepts necessitate new approaches to assessment design. This paper documents the iterative design of an assessment task intended to trace the crosscutting concept of energy across three different disciplinary contexts in high school science. We review and synthesize literature on performance tasks, three-dimensional task design, and research into student thinking about energy. Then, working with examples from a research-practice partnership, we identify three tensions that have emerged in the process of design: tension between practice-as-embodied in the task and the current state of practice in our partner district, tensions in creating scorable outcome space as students model and explain energy across systems, and tension in asking students to represent energy in ways that pose challenges for disciplinary experts. We close by summarizing ongoing challenges for assessment designers engaged in designing assessments for crosscutting concepts.

Introduction

The introduction of the Framework for the Next Generation Science Standards [NGSS] changed the US vision of science teaching and learning from a two-dimensional, topics and skill-based approach to a three-dimensional approach. The new science teaching and learning standards foster thinking in the three dimensions: science and engineering practices, core ideas and crosscutting concepts (National Research Council [NRC], 2012). The science and engineering practices subsume cognitive, social and physical practices that are required to investigate and build theories and models about the natural world. The core ideas are key organizing concepts of single disciplines. These two dimensions of the NGSS are similar to those that have previously been documented as part of international inquiry-oriented science teaching reforms (e.g. OECD, 2017).

However, the crosscutting concepts in the NGSS are a new way of thinking about ideas that span all of the sciences, as they are of broad importance and have applications in all science domains (NRC, 2012). These concepts, which include systems and systems thinking, patterns, and cause and effect, can be applied to a wide range of phenomena, and are included in every one of the individual NGSS performance expectations. The crosscutting concepts necessitate not only new approaches to science teaching that help students to foreground and connect these overarching themes, but also new, multicomponent three-dimensional assessments which will be able to evaluate the teaching and learning of all three dimensions in the NGSS (NRC, 2014).

In this paper, we identify tensions that surfaced as we designed a three-dimensional performance task that foregrounded the crosscutting concept of energy in the context of a long-term, mutualistic collaboration between our research team and a large culturally and linguistically diverse school district. The task focuses on the disciplinary core ideas of “cycles and energy transfer in ecosystems”, “chemical reactions” and “forces and motion”, the scientific practice of “developing and using models” and “constructing explanations”, and the crosscutting concept of “energy” (NRC, 2012). Drawing on data from our design of this task, iterative rounds of feedback with teachers, science curriculum coordinators, scientists, and students, we identify challenges facing curriculum and assessment designers as they move into the space of three-dimensional assessment.

Theoretical and conceptual foundations

Consistent with what Ford & Forman (2006) called the ‘practice turn’ in sociocultural theory, and the following focus on engagement in practice as a goal for disciplinary learning, the Framework for the Next Generation Science Standards (NRC, 2012) foregrounds engagement in practice for students as they learn disciplinary core ideas and apply crosscutting concepts. This shift in focus seeks to move the field of science assessment away from a focus on knowledge alone, as has been the traditionally privileged outcome of educational contexts for decades, and toward a definition of learning as changes in participation in disciplinary practices over time (e.g. Wenger, 1998). This shift in the way we theorize about what students do in assessment contexts also repositions the way we think about knowledge. From a sociocultural perspective, assessment designers are no longer considering knowledge as the only outcome (e.g. Shepard, 2000), but rather focus on the ways that students engage in practices as they demonstrate their disciplinary knowledge.
The consequences for the design of assessments are clear: assessments can no longer focus only on eliciting the different types of knowledge that students bring to different contexts, but must also create opportunities for students to engage in scientific practices (NRC, 2014). While this perspective is not new, its primary emphasis on multiple dimensions of science learning is. Many performance assessments developed in the 1990’s prioritized students’ engagement with concrete materials as they solved contextualized problems (Solano-Flores & Shavelson, 1997) and the ways in which students engaged in processes of inquiry as an outcome of their science learning. However, the extent to which these tasks actually engaged students in higher-level cognitive processes has been questioned (Baxter & Glaser, 1998). For example, the ‘Paper Towels’ task assessed students’ experimental design of which brand of paper towels absorbed the most water (Baxter and Shavelson, 1994), without a focus on underlying scientific principles.

In the current reform context, science assessment design frameworks draw deeply upon these previous efforts (e.g. Pellegrino, Chudowsky & Glaser, 2001), but are also informed by new perspectives on student engagement in scientific practices, such as modeling (e.g. Schwarz et al., 2009; Windschitl, Thompson & Braaten, 2008), explanation (e.g. McNeill et al., 2006), and argumentation (e.g. Bricker & Bell, 2008). In addition, three-dimensional science assessments seek to create opportunities for students to demonstrate their learning in the context of compelling, real-world phenomena (NRC, 2014), as is the case with some of the sample tasks that have been generated recently (Achieve, Inc., 2014). In this sense, the design of assessment tasks is repositioned from the ways we have previously thought about transfer of learning (NRC, 2007) to instead focus on contextualized phenomena aligned with students’ interest (NRC, 2014).

Multicomponent assessment design

The new vision for assessment tasks based on the Framework challenges assessment designers to use multiple inter-related questions or components to fully assess the performance expectations included in the NGSS (NRC, 2014). These tasks will be developed following evidence-centered or construct-centered design processes (NRC, 2014), and will involve an iterative process composed of multiple steps: analyzing and detailing the cognitive domain to be assessed (others have called this ‘unpacking,’ see Stevens, Delgado & Krajcik, 2010), identifying the inferences that the assessment is designed to support about student learning and determining the types of evidence necessary to support those inferences (Pellegrino et al., 2001), designing tasks that will collect that evidence, and determining how to model evidence to support valid conclusions (NRC, 2014).

Given that the Framework (NRC, 2012) vision is still new, and the design processes for Framework-aligned tasks even newer (NRC, 2014), examples of what these assessment tasks look like in practice are only beginning to emerge. Achieve, Inc. has released sets of sample tasks (Achieve, Inc., 2014), all of which may take days, even weeks to complete. Pages of single-spaced task prompts are followed by multiple diagrams, data, and images for students to analyze, leading to long tasks that would require significant tailoring for use in teachers’ school contexts, as well as scaffolding to support students in responding to the tasks. Clearly, the field is still developing images of what form this type of assessment will take.

Criteria and constraints for the design of a three-dimensional task

Our efforts in this domain have taken place in the context of a larger research-practice partnership (Penuel et al., 2011) intended to develop a system of three-dimensional classroom assessments. This partnership, which dates to 2014, began at the initiation of the school district, which reached out to researchers at our University for support around NGSS-aligned formative assessment design. This mutualistic collaboration (Coburn & Penuel, 2013) involves long-term commitments from researchers with deep support from district administration in our partner district, located outside a large city in the Western US.

Since that time, three externally funded grants have supported our partnership as we have developed a series of multicomponent, pre-post assessments to model student learning within and across school years, creating opportunities for longitudinal tracking of cohorts of students as they move through high school physics, chemistry, and biology. While our initial assessment design efforts spanned multiple disciplinary core ideas and crosscutting concepts, we have simplified our work by focusing on the scientific practice of modeling (Passmore & Svoboda, 2012; Passmore, Schwarz & Mankowski, 2017), and then tracing energy as both a disciplinary core idea unifying instruction across high school physics, chemistry, and biology, as well as a crosscutting concept across these disciplines (c.f. Park & Liu, 2016). We describe this in greater detail in the following section.

Energy: Crosscutting concept and disciplinary core idea

Energy occupies a unique position in the Next Generation Science Standards, as it is both a core idea across the different science disciplines, as well as a cross-cutting concept. Studies in different scientific disciplines have investigated both implicit and explicit learning of the concept (e.g. Park & Liu, 2016; Opitz et al., 2015; Opitz,
To date, four main characteristics of energy have been identified: energy is present in different forms; energy can be transformed from one form to another or transferred without changing its form; energy is degraded, whenever it is transformed; and the overall quantity of energy is conserved (Duit, 1984). These characteristics are more or less prominent in disciplinary topics. For example, in life sciences, teaching of the energy concept mostly focuses on energy transfer and transformation processes in open systems (Opitz et al., 2016), whereas in physics all four characteristics are introduced in middle school and are revised with quantitative considerations in high school (Neumann et al., 2013).

Learning about the four characteristics of energy is the subject of many studies (e.g. Neumann et al., 2013, Jin & Anderson, 2012; Nordine, Krajcik, & Fortus, 2011). For example, Jin and Anderson (2012) identified a hierarchical structure of energy understanding for biology students, and Opitz and colleagues (2016) described the energy conceptions of biology students in middle school. Although students' understanding of energy forms and transfer/transformation have been shown to increase over time, students also maintain many prior ideas they held before entering school after instruction (Jin & Anderson, 2012; Lancor, 2014).

Students’ conceptions of energy in physics or chemistry develop in a similar way and student learning also seems to be hierarchical. In physics, learning about transfer and transformation is associated with degradation (Neumann et al, 2013) and in chemistry energy transfer and transformation is associated with forms (Teichert & Stacey, 2002). The few studies that have assessed energy across all disciplines (e.g., Opitz, et al., 2017; Park & Liu, 2016) have found high latent intercorrelations between the energy understanding in different disciplines. These findings indicate, that there is little variance in student learning between disciplines while maintaining a large variance within each discipline (Park & Liu, 2016).

**Research Questions**

The studies reviewed above set a key challenge for three-dimensional science assessment: to determine the types of tasks that will be able to capture the development of student engagement in scientific practices and disciplinary core ideas as they span across crosscutting concepts that students encounter in multiple years of study, such as energy. Specifically, we have proceeded with the process of designing an assessment task in the context of our research-practice partnership, and seek in this paper to respond to the following research questions: How can we develop a three-dimensional task to assess the crosscutting concept of Energy? What tensions and challenges emerge in this process? In responding to these questions, we seek not only to identify tensions and the ways we addressed them in our study, but also to inform future assessment design in this area.

**Method**

Our paper, and the larger project in which it is embedded, uses a Design-Based Implementation Research (DBIR) approach in the context of a research-practice partnership (Coburn & Penuel, 2016) to develop and test innovations fostering alignment and coordination to improve classroom practices (Penuel et al., 2011). We conduct rapid cycles of design that allow us to negotiate means and goals across multiple stakeholders in real time (Cobb et. al, 2013); this paper provides a case analysis of several cycles of rapid prototyping a single task.

**Task design procedure**

Our task design followed a multi-step, iterative approach (NRC, 2014). We started by identifying the NGSS performance expectations associated with energy as disciplinary core ideas and crosscutting concepts. Next, we ‘unpacked’ the ways energy was discussed in these performance expectations by building on and expanding frameworks for energy in physics, following Neumann et al. (2014)’s learning progression for energy forms, transfer/transformation, conservation, and degradation/dissipation. Next, working from similar assessments tracing energy across multiple frames of reference (e.g. Ambitious Science Teaching, 2017; Neumann, Fortus & Nordine, 2017), we developed separate versions of the task for every science discipline. This first task draft was piloted with high school environmental science students in our partner district (N = 26).

Based on students’ response patterns, as well as the desire to move closer toward a task that could be used at any grade level or disciplinary focus in high schools, we used the students’ pilot data to guide our revision of the three separate tasks and combine them into a single crosscutting format. This second version of the task was piloted with disciplinary experts, scientists in the domains of physics, chemistry and biology (N=5). All of the scientists were asked to solve the task and two of the scientists were interviewed about their responses. Additional feedback about the clarity and accuracy of the task was used to inform the next iteration. At this phase we also developed pilot versions of a rubric to score the task, using scientist expert responses to populate the top levels of the rubric, and information from the SOLO Taxonomy (Biggs, 1979), Park and Liu’s (2016) study of energy as a crosscutting concept, and Neumann et al.’s (2013) five ‘big ideas’ about energy. At this same phase, we shared the task with communities of science teachers in the partnership (4 teacher groups,
16 teachers) and collected information about their responses and reactions in detailed fieldnotes. Teachers discussed the task in the context of their own understanding of the content as well as in the context of their perception of their students’ understandings.

The third iteration of the task was piloted with university undergraduates in freshman-level chemistry courses, as they are representative of students who have completed three years of high school science (N=23). Following this administration, the research team conducted focus sessions (Briggs & Peck, 2015) with representative samples of student responses to identify further revisions to the task, as well as to the scoring rubric. Based on these experiences, we developed a revised, simplified format for the task.

Sources of data and analytic approach
We draw on several sources of data collected across multiple settings during the spring and fall of 2017, including in-depth running design notes from weekly university-based research team meetings, handwritten fieldnotes made during bi-monthly design meetings with district partners, meeting agendas, artifacts, and fieldnotes from school-based teacher learning community meetings, facilitation guides, fieldnotes, and artifacts created at all-district professional development meetings, student responses to pilot versions of the task, and fieldnotes and artifacts from administering the task to scientists.

The authors of this paper met multiple times to review and discuss the data, and identified initial tensions. These tensions were shared with members of the research team, who then interrogated our developing ideas. We then tested the initial tensions we identified against other forms of data, refining them as we developed the case study. Our identification of tensions emerging from across these multiple settings and sources of data occurred during these conversations, as well as in the course of our regular research team activities. After we wrote up our initial emergent tensions and claims to support them within the research team, we shared those claims with district science coordinators, scientists, other members of our research team, as well as other learning scientist colleagues. We integrated their feedback and reflections in the final draft of this paper.

Emergent tensions in process of design
Our analysis of the preceding sources of data have led us to identify three emergent tensions surfaced in our attempts to design a three-dimensional assessment task intended for use in tracing the development of students’ understanding of how to model energy in systems across multiple school years. We describe each of these tensions below, with illustrative examples of our iterative cycles of design, piloting, and revision.

Tension between practice-as-embodied in the task and current classroom practice
The original idea for the energy task came from Sabrina and Liz, the two district science coordinators in our partner district, while examining published examples of assessment in Nordine (2017). Noting that using contexts related to sustainability was a priority in their district, Sabrina seized upon the idea of extending the example of biofuels into a context that might be used across high school physics, chemistry, and biology. We began developing a modeling and explanation task that would allow students to trace energy as a crosscutting concept across systems that represented different disciplinary core ideas in the three science domains.

As we mocked up versions of the task, Sabrina and Liz reflected on whether or not it was the kind of assessment that would align with current science teaching practice in the district, which largely involve traditional instruction representative of the majority of US science classrooms (e.g. Banilower et al., 2012). District leadership had recently asked her to justify her department’s focus on the three-dimensional vision of learning, even though she was working in a non-NGSS state. Noting that the classroom practice of most teachers in the district was nowhere near the three-dimensional approach we were aiming for, Sabrina exclaimed, “I feel like I’m on Pluto.” When asked to explain more about what she meant, she elaborated that she felt like the vision we were going for in our partnership was so distant from what was happening in classrooms in the district that the design for the assessment task was starting to feel like outer space. Nevertheless, both Sabrina and Liz committed to the task as ‘aspirational’ to inform vision for science teaching and learning in the district, and they intend to begin using this task as a way of compelling changes in classroom practice. However, in the meantime, this also means that piloting and initial use of the assessment is taking place in classrooms in which students have little experience with or opportunity to learn through modeling and explanation; that is, practice-as-embodied in the task feels, at times, billions of miles away from what students experience on a daily basis.

Tensions between modeling, explanation, and scorable student responses
Long-standing lines of research in science education have examined the ways in which students’ abilities to create and use models support their learning of important science concepts (e.g. Schwarz et al., 2009), and studies of the ways that students create and revise models have identified multiple types of scaffolds, such as
checklists, that help students engage in this scientific practice (Kang et al., 2014). In parallel, researchers have also established frameworks and scaffolds for engaging students in the scientific practice of constructing explanations (e.g. McNeill et al., 2006; Songer & Gotwals, 2012). In our work, we have sought to create a task prompt that builds on lines of research from both of these traditions to move toward engaging students in using their models to create explanations, in which students create a model for a given phenomenon, and then use that model to develop an explanation of a specific phenomenon.

The left side of Figure 1 shows the first version of the task that had a large, blank space for students to draw a model of how algae produces oil that could be used to fuel a bus (lower-left frame), as well as an unstructured space for students to write an explanation for how energy flows in this system. We provided images at the upper-right and lower-left to scaffold students’ responses about energy transfer from the sun to the algae, as well as energy helping the bus move. We created similar versions that alternated the focus at the center, one with a larger frame for biology students (focusing on how algae capture energy from the sun through photosynthesis) and another for physics students (focusing on how energy helps a bus move up a hill). Our initial pilot with high school students quickly indicated that students were unsure what to make of the different reference frames, as well as the large blank box at the center, and wrote little to no explanation. Students were unfamiliar with the phenomenon of using algae to produce biofuels and were unsure how to approach the task.

At the same time, as a research team, we engaged in conversations about difficulties we would encounter in modeling student understanding of a crosscutting concept if the task was so intimidating as to provide no place for students to begin. We were also concerned that students would not explain the entire process in a large outcome space, and reflected that breaking the explanation into smaller pieces linked to the different pieces of the model might help us generate more scorable information. We also had concerns that the three versions of the task might be non-comparable across grades, creating difficulties in tracking students across multiple years of science courses.

Our solution to these design challenges is shown in the right side of Figure 1, a portrait-oriented version of the task that repositions the question in the context of corn, a familiar crop to the students in our partner district, and ethanol, a substance students commonly see or hear about at local gas stations. It breaks the modeling of energy transfer and transformation into four, equally-sized boxes (corn plants capturing energy from the sun; distilling ethanol from fermented glucose; combusting ethanol in a piston; a bus moving on a road). Each box is then matched with a specific question about energy in that part of the model, with separate outcome spaces. This version of the task, then, was intended to strike a balance between having students create a model and use that model to create an explanation, as well as having smaller pieces of the task with more accessible outcome spaces for students.

In this process of designing the task, as Figure 1 shows, we also experimented with different checklists for vocabulary, modeling and explanation. Following Kang and colleagues’ (2014) findings, we knew that providing some level of scaffolding would create more opportunities for students to make their thinking visible in the task, and would increase the quality of their models (e.g. focusing on both visible and invisible processes).
and explanations (explaining transfer and transformation in the different parts of the model). We shared this version of the task with our teacher partners, collecting both their responses and their impressions of the task.

When performing initial modeling work to explore the ways students were likely to think about this crosscutting concept, our team developed a task, based on Eisenkraft (2017), in which we built an inefficient calorimeter and then created initial and revised models of how energy would flow through the system when we burned a piece of Pirate Booty. The first time we completed the task, members of the research team with biology, environmental science, and physics backgrounds all included different types of energy in their models. The physics major wrote about kinetic and potential energy, common ways of talking about energy in his discipline, whereas the biologist and environmental scientists focused on energy transfer and transformation within the system, consistent with their experiences modeling energy and matter flow in ecosystems. Our experiences engaging teachers at our partner schools in this calorimeter activity yielded similar results.

When we gave an updated version of the task shown in Figure 2 (left) to disciplinary experts, we were seeking to prompt them to not only focus on how energy flowed through the system, but also to make macro- and micro-level connections (e.g., not only noting that energy transfer and transformation are occurring in photosynthesis, but also writing out equations for carbon fixing in the process of photosynthesis). We were surprised to find that most of the experts took on the sections associated with their field first and later apologized for their lack of familiarity with other sections. For example, the physicist stated that “... the complexity of the photosynthetic process is outside of my specialty,” and the biologist noted on her task, “I found this difficult because I don’t teach this topic.” Counter to our expectations, these experts - all of whom had extensive knowledge of contexts inside and outside of their fields - were also feeling limited by the same disciplinary boundaries uncovered by members of the research team. This led us to wonder about the ambition of the crosscutting concepts themselves, since they were seeking to represent larger ideas in science that even pushed the boundaries of the ways scientists think on a daily basis.

As we piloted the task with high school and college-level students, we also became increasingly aware that the boxes around the different elements of the model - vestiges of the original three-version task - actually might be reinforcing these disciplinary boundaries that were challenging for the scientists. As such, the later versions of the task removed both the boxes around the different parts of the model with specific disciplinary foci, as well as the suggested ‘zoom-out’ boxes intended to prompts students to draw micro-level processes.

These experiences prompted us to simplify the task to allow a broader aperture of responses, where disciplinary experts might be able to ‘go deep’ at the micro- or nano-scale, while also making the task accessible to students responding on the macro-scale on the basis of their everyday experiences. We also hoped that we could find a task format that would allow students to work with the ideas of the crosscutting concept of energy without necessarily being turned off by checklists of vocabulary words with which they might not be familiar. Thus the revised version of the task, shown on the right side of Figure 2, included fewer scaffolds for both the model and the explanation outcome space. At the time of publication of these proceedings, this version was being administered to physics, chemistry, and biology students in our partner district.

Discussion
As the field moves toward new ways of thinking about science learning, new methods for developing classroom-based assessments of this learning are necessitated (NRC, 2014). Our experiences developing the Energy Assessment Task has illustrated that a design in ‘outer space’ may seem that way not only to the
students and teachers in our partner district, but also to scientists whose daily work takes place within disciplinary constraints that reinforce the very boundaries that new ways of thinking about science learning - in particular, crosscutting concepts - are intended to diminish. The tensions we have identified are likely only the beginning of those that assessment designers and those working in partnership with districts, schools and teachers are likely to uncover. However, we emphasize that such aspirational assessments are an important component in the new systems of curriculum materials and professional learning experiences currently being developed to support Framework-aligned learning experiences for students (e.g. Reiser et al., 2017).

Returning to the situated perspective that we bring to this work (Greeno, 2006), we acknowledge the critical role that tools such as tasks like these might play not only in reorganizing participation structures in the classrooms we support, but also to create opportunities to discuss the ways in which the task embodies a shared vision for the district as it moves toward different ways of thinking about science learning outcomes (Wenger, 1998). We also acknowledge the number of critical questions that we direct to those developing assessments in this domain, including: What does it mean to move toward a vision of assessment that is so ‘out there’ that even scientists are challenged by thinking in that way? How can we scaffold student participation in assessments like these when students’ opportunities to learn through instruction aligned with the task are still limited? What does this mean for the vision for learning and teaching in the NGSS?

References


Acknowledgement
This material is based upon work supported by the National Science Foundation under Grant No. 1561751. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Investigating the Impact of an Online Collaboration Course on Students’ Attitudes and Learning

Soniya Gadgil, Maggie Braun, Martha Harty, Ken Hovis, and Marsha Lovett
soniyag@andrew.cmu.edu, mabraun@andrew.cmu.edu, mh51@andrew.cmu.edu, khovis@cmu.edu, lovett@cmu.edu
Carnegie Mellon University

Abstract: Collaboration and teamwork are critical skills in today’s workplace, but teaching them effectively remains a challenge. Collaborative U, an online instructional resource that students complete at the start of a major group project, was designed to address this issue. We present results from a randomized controlled trial comparing Collaborative U to a control condition equated on format and duration. Students in the Collaborative U condition maintained their initial high positive attitudes towards collaboration from pre to post, whereas students in the control condition showed a decline. Furthermore, students who completed Collaborative U learned more from pre to post and also rated team process and final products significantly higher than those in the control condition on a variety of attributes. Our findings offer encouraging evidence for the efficacy of a short, online intervention in supporting teamwork skills, and have important implications for integrating team-based activities into classrooms.

Keywords: collaboration, teamwork, higher education, open learning, online learning

Introduction
The benefits of learning with one or more peers have been documented extensively in the learning sciences and educational literature (Johnson, Johnson, & Smith, 1987; Roschelle & Teasley, 1999). Learning collaboratively has been shown to benefit students of various age groups, and institutions of higher learning have been increasingly integrating collaboration into their classrooms as a way to increase active learning (e.g., Barkley, Cross, & Major, 2014). Engaging in team projects during college years also helps students develop important skills for today’s workplace, where most will be required to work in teams, as emphasized by several frameworks that outline key 21st century skills (AAC&U, 2015; Dede, 2006).

Despite the myriad cognitive and motivational benefits of collaboration, it does not always yield positive outcomes. For example, teams may fall prey to the effects of social loafing (Karau & Williams, 1993), fear of evaluation (Paulus & Dzindolet, 1993), or production-blocking (Diehl & Strobe, 1987), which may hinder a team’s productivity. Students sometimes report dissatisfaction with project teams (Barfield, 2009), and worry about the fairness of their individual grades (Comer, 1995), and instructors may grow weary of troubleshooting collaborative activities in the classroom (Panitz & Panitz, 1998).

In this research, we report results from an online intervention designed to provide students with the tools to effectively engage in face-to-face collaborative interactions, such as working together on a team project. The goals of the intervention were to improve the quality of students’ collaboration by helping them engage in productive team skills and contribute evenly to a common goal. We describe results from a large-scale randomized controlled trial, in which first-year college students were assigned by section to an intervention group that participated in online instruction focused on improving collaboration skills, or to a control group that received an online instruction of similar format and duration but not related to collaboration. We investigated the effects of the intervention on students’ performance, attitudes, and learning.

Background and literature review
The term ‘collaborative learning’ refers to the practice of assigning students to work in teams of two or more to produce or create something (Dillenbourg, 1999). It is incorporated into college classrooms in a variety of ways, including but not limited to group projects, think-pair-share activities, and peer reviews, among others. However, the degree to which groups are successful differs. Many studies in higher education have found collaboration to be better than individual learning (see Lou, Abrami, & d’Apollonia, 2001; Springer, Stanne, & Donovan, 1999 for meta-analyses), but others have found no difference (e.g., Crooks, Klein, Savenye, & Leader, 1998), and a handful have found collaboration to be worse than individual learning (see Kirschner, Paas, & Kirschner, 2009 for a review).
Several studies have found that when students receive collaboration support, it improves their collaboration, and leads to better outcomes compared to when they collaborate in the absence of support (Weinberger, Stegmann, & Fischer, 2010). Collaboration support may take the form of collaboration scripts (Rummel & Spada, 2005), prompts (Xie & Bradshaw, 2008), guided activities, collaborative argumentation (Asterhan & Schwartz, 2009), workshops, or readings that address collaboration. Some of these strategies are strongly supported by evidence (e.g., Rummel and Spada, 2005), whereas others have not received strong support (e.g., Rau, Kennedy, Oxtoby, Bollom, & Moore, 2017). In the present work, we took a cognitive task analysis approach (see Crandall and Hoffman, 2013 for a review) to identify, instruct, and provide practice on several key skills that team members need. The idea was to promote learning of key skills in the online resource in a variety of contexts, some designed to be fairly realistic, and immersive experiences, so that students would be able to transfer these key skills to their real-world teams.

Learning interventions tested in controlled laboratory settings often show promising results, but when applied at scale in larger classroom environments, the effects are frequently not sustained (Elmore, 1996). It is also difficult to assess the efficacy of such interventions at scale, because implementing randomized controlled trials in educational settings presents numerous challenges (Cook, 2002). As a result, evidence on efficacy of learning interventions often falls short of the “gold standard” of randomized controlled trials (Whitehurst, 2003). In this research, we aimed to apply findings from basic research and test our intervention at scale, in the context of a “gold standard” randomized controlled trial.

**Present study**

We created an online instructional module that teaches collaboration skills as a form of collaboration support. We tested its effectiveness thoroughly by delivering it within a randomized controlled experiment, such that one group of students received our intervention, and the other received a control module of similar duration and format. Students applied those skills in a face-to-face collaborative project that they completed as part of a general education seminar taken by all first-year math and science students at Carnegie Mellon University.

**Method**

**Course description**

EUREKA! is a seminar-based course that is taken by all first-year students in the Mellon College of Science and is a requirement for graduation. During the seminar, students are exposed to campus resources that promote academic and personal success, and they participate in recurring self-assessments that promote personal well-being, academic improvement, and ethical decision making. Most pertinent to the current study, students also work on team projects focused on a scientific research topic of their own choosing. Students work over much of the semester in teams of 3-5 to produce a short video on a research topic of their choosing, and lead an in-class discussion/activity as a group. All students took an individual pretest before the online intervention, to assess their personal attitudes towards collaboration and their knowledge of good collaboration practices. An identical immediate posttest followed the intervention (also completed individually). At two points during the group projects, students rated their peers as well as themselves on contributions to the team. Finally, students were evaluated on their final products by both their instructors and their peers (see Figure 1 for timeline).

![Timeline of Activities](image)

**Developing the intervention: Collaborative U**

In 2015, Carnegie Mellon University made acquiring collaboration and teamwork skills a critical priority for all students and a goal of its strategic plan. An interdisciplinary team was assembled around a proposal to build
evidence-based, scalable, faculty-friendly basic training for students involved in group projects. The resulting prototype, Collaborative U, is a 3-hour online instructional module informed by rigorous research and practical experience on structuring effective collaborative teams. For example, Woolley and colleagues have shown that social sensitivity and equal participation among members are some of the key factors that improve effectiveness of teams (Woolley, Aggarwal & Malone, 2015). We also leveraged interactive training methods developed and refined over a decade of training student project teams in traditional face-to-face workshops by collaboration and conflict skills trainers.

Figure 2. Screenshot showing one of the “learning by doing” activities in Collaborative U.

Collaborative U combines online modules that students work through individually with face-to-face practice activities performed as teams. Instructors assigned Collaborative U as part of EUREKA! at the start of the team projects. Students completed two hour-long units individually online, covering basic diversity communication and conflict resolution skills. After each unit, students completed a 20-minute activity in their project teams, discussing members’ strengths and differences and practicing skills for navigating conflicts constructively. The online modules involved an extended video-based interactive scenario for skill demonstration and practice, and a variety of other realistic practice environments, such as texting and video chats (see Figures 2 and 3 for examples).

Collaborative U was deployed using the Open Learning Initiative (OLI) platform. OLI is an open educational resources project at Carnegie Mellon University that allows instructors to develop online courses consisting of interactive activities and diverse multimedia content. OLI courses are sometimes delivered asynchronously without an instructor; in other cases, they are used by instructors to support and complement face-to-face classroom instruction. Studies that compared student learning in a blended (face-to-face plus OLI) course on introductory statistics to a traditional, face-to-face version of the same course showed that students learned better and in half the time in the blended-OLI format (Lovett, Meyer, & Thille, 2008; 2010).

Figure 3: Example of an immersive, real-world activity in Collaborative U.

Control condition
Like Collaborative U, the control condition was a set of OLI modules, with a similar format but focused on a different topic. The control modules addressed visual design, so there was a reasonable mapping of that instruction to help students make better team products. The in-class activities that teams worked through after the individual unit work were focused on visual design for all students in the control group.

Study design
We investigated the impact of Collaborative U on actual team function in a course assignment. The EUREKA! course had thirteen recitation sections, seven of which were randomly assigned to receive the Collaborative U OLI modules, and the remaining six to receive a control OLI resource targeting visual design. Students from both conditions completed pre/post-tests on teamwork. The time on task was roughly equivalent across groups. Two 20-minute face-to-face team activities and discussions were completed in class (focused on either teamwork or visual design, respectively) so that students could apply what they learned in the OLI modules.

Participants
Participants were first-year students enrolled in the EUREKA! seminar at Carnegie Mellon University. 120 students completed the Collaborative U module, and 106 completed the control module. The gender distribution across both conditions was 50-50.

Research questions
We assessed the impact of collaborative support in the form of an online instructional module, Collaborative U, relative to a control module of equal duration on students’ team process, learning, and performance. Specifically, we asked the following questions:

1. How does Collaborative U impact students’ attitudes towards collaboration, and their knowledge of good collaboration practices from pre to post, compared to control?
2. How does Collaborative U impact student’s self-ratings and ratings of their collaborative peers in terms of team process, relative to control?
3. How does Collaborative U impact the final team product, in this case, a project video and in-class discussion, compared to control?

Measures

Pre and post tests
Before and after the intervention, students responded to two five-item questionnaires, on which each item was rated on a five-point scale, with 1 corresponding to “Strongly Disagree” and 5 corresponding to “Strongly Agree”. These scales measured students’ attitudes towards collaboration and their knowledge of what makes a good team. They also took a conceptual knowledge test on collaboration before and after completing the intervention on which they could score between 0 and 4 points.

Students’ self ratings and peer ratings of team process
Students rated themselves and their teammates at two time points on the following attributes: promptness and attendance at team meetings, preparedness at meetings, effort level, attitude, helpfulness, content knowledge, effectiveness toward project goals, flexibility, and desirability as a team member. The ratings were on a scale of 1-4, with 1 indicating never/rarely, and 4 corresponding to always. The first peer rating was completed at week 11, about 5 weeks into the team project, immediately after the video was due, and the second peer rating was completed at week 13, after the teams led in-class discussions.

Peer reviews of final project videos
Each student used a rubric to grade videos produced by every other team. They rated the videos on a scale of 1-10 on the following attributes – objective, summarization, clarity, time distribution, interest, audio quality, video quality, group dynamics, and relevance, and assigned an overall score.

Results
In this section, we will first present results from pre and post assessments for the Attitudes towards Collaboration and Characteristics of good team scales, as well as the content knowledge test. We will then describe results from students’ self-ratings and peer ratings of team processes. Finally, we will present results from peer reviews of final project videos.
Pre and post tests
Overall, on students’ mean rating on the Personal attitudes towards collaboration scale, there was a marginal effect of test time, such that both conditions changed significantly from pretest to posttest, $F(1, 208) = 2.81, p = .09$. Students in both Collaborative U and control conditions showed a marginal decline in their mean scores from pre to post. However, this decline was driven primarily by the control condition (see Figure 4a). Analyses of simple effects indicated that the decline from pre to post was marginally significant for the control condition, $F(1,208) = 3.68, p = .056$.

On the Characteristics of good teams scale (see Figure 4b), there was no main effect of test-time, $F(1, 208) = 0.39, ns$, indicating that overall, there was no change from pre to post. The interaction between test-time and condition was not significant $F(1,208) = 1.14, ns$. Examination of means suggests that students in Collaborative U showed a slight improvement, and those in the control condition showed a slight decline, however, since the main effects and interaction did not reach significance, follow up tests were not conducted.

Students also took a conceptual knowledge test on collaboration before and after completing the intervention, on which they could score between 0 and 4 points. Figure 5 shows the mean pretest and posttest scores for each condition. A repeated measures ANOVA was significant for test-time, $F(1,209) = 36.47, p = 0.00$, indicating that all students improved from pre to post. The main effect for condition was also significant, indicating that the means for the Collaborative U and control conditions were significantly different, $F(1, 209) = 4.08, p = 0.04$. A test-time by condition interaction was also significant, $F(1,209) = 32.58, p = 0.00$. Follow up tests indicate that the difference between pretest and posttest was significant for the Collaborative U condition, $F(1,209) = 78.26, p = 0.00$, but not for the control condition, $F(1,209) = 0.049, ns$.

First peer review and self review
On the first peer review that took place in week 11, students in the Collaborative U condition rated their peers higher than those in the control condition on a variety of attributes (see Figure 6). For effort level $t(536) = 3.099, p = .002$, attitude $t(533) = 2.765, p = .006$, helpfulness $t(536) = 2.182, p = .029$, flexibility $t(536) = 3.65, p = .0002$, and desirability as a team member $t(536) = 2.91, p = .003$, ratings by students in Collaborative U condition were significantly higher compared to those in the control condition. The overall rating for team members was also significantly higher in Collaborative U compared to the control condition $t(536) = 2.03, p = .04$. On content knowledge $t(536) = 1.66, p = .09$, and effectiveness towards project goals $t(536) = 1.73, p = .08$, they were
marginally higher. On two other measures — attendance and preparedness at meetings, the two conditions were not significantly different. Interestingly, on self-ratings, students in Collaborative U and control were not significantly different on any of the attributes.

### Second peer review and self review

On the second peer review that took place in week 14, students in Collaborative U continued to rate their peers higher compared to those in the control condition on several measures. For attitude ($t(475) = 1.67, p = .09$), content knowledge ($t(476) = 1.90, p = .058$), and flexibility the difference in rating was marginally significant ($t(477) = 1.85, p = .06$), favoring Collaborative U. On ‘desirability as a team member’ the difference was significant ($t(475) = 3.17, p = .001$, again favoring Collaborative U (see Figure 7). On preparedness at meetings, effort level, helpfulness, effectiveness towards project goals, as well as overall rating, the differences were not significant. Just as on the first round of ratings, students in Collaborative U and control did not differ significantly on self-ratings for any of the attributes.

### Peer Reviews of Team Products

On ratings of team products demonstrated via team-led, in-class discussions, students in Collaborative U received higher ratings compared to those in the control condition on several measures. On objective ($t(842) = 2.05, p = .04$), summarization ($t(784) = 2.30, p = .02$), clarity ($t(785) = 2.86, p = .004$), time distribution ($t(783) = 3.40, p = .0007$), group dynamics ($t(784) = 3.68, p = .0002$), relevance ($t(762) = 2.49, p = .01$), and overall score ($t(843) = 2.41, p = 0.016$, the differences were significant, favoring Collaborative U (See Figure 8). On interest ($t (784) = 1.68, p = .09$), the difference was marginal. Finally, and of note, on audio quality and video quality — attributes not directly related to collaboration, the two conditions were not significantly different.

---

**Figure 6. First Peer Rating.**

**Figure 7. Second Peer Rating.**
In this paper, we describe results from a randomized controlled trial comparing a short online instructional module for collaboration support with a control condition equated on format and duration. We found compelling evidence that completing our instructional intervention had significant impacts on students’ team processes and products. On personal attitudes towards collaboration, students in Collaborative U maintained their initial high scores from pre to post, whereas the control condition showed a significant decline. This result suggests that engaging with the Collaborative U modules provided an inoculation effect that buffered against the decline experienced by students in the control condition. Students in Collaborative U also learned more content knowledge from pre to post. Further, when compared with a control condition, students in Collaborative U rated their peers higher on several attributes critical to good team participation. Students in Collaborative U also generated better final products as a result of the team activity, as measured by peer review ratings of various attributes.

While our findings are encouraging, we note a few caveats. First, our study was done within the context of a highly selective private university. In order to be more generalizable, our study needs to be replicated across different settings, such as public universities and community colleges. Second, we used peer reviews to rate teammates’ contributions to teams. However, more objective measures of quality of collaboration are available, such as the Collective Intelligence Battery (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010) which would be useful in corroborating the results of peer reviews and project grades. Third, while our study showed an advantage for the Collaborative U condition on immediate assessments, we need to understand how Collaborative U affects student attitudes and perceptions of collaboration in the long term, and whether it helps prepare them for future collaboration opportunities. We are currently pursuing these questions by longitudinally following up a subset of the students who participated in the present study. These robust measures of learning will further solidify the evidence that online collaboration support can lead to better learning and collaborative performance.

Despite the limitations discussed above, our findings show strong evidence that collaboration support in the form of an instructional module can improve students’ team performance and their knowledge and attitudes toward collaboration. By comparing it against a time-equated, relevant control condition, we showed that it was not simply additional instructional time that explains the superior outcomes for the Collaborative U group. Given the increased use of collaboration as a form of active learning in higher education, we believe that our findings have wide applicability. Delivering the intervention through an online platform makes it easier to implement at scale.

References


Dillenbourg, P. (1999). What do you mean by collaborative learning?.


**Acknowledgments**

This research was approved by the CMU Institutional Review Board (IRB) under the protocol “Study2016_00000291: Collecting Data on the Impacts of Educational Interventions”. This work was facilitated by the Teaching as Research service, provided by CMU’s Eberly Center for Teaching Excellence and Educational Innovation and was supported by a Carnegie Mellon University ProSEED/ Simon Initiative Seed Grant to M. Harty & R. Vituccio.
How Broad is Computational Thinking?
A Longitudinal Study of Practices Shaping Learning in Computer Science

Chris Proctor, Stanford University, cproctor@stanford.edu
Paulo Blikstein, Stanford University, paulob@stanford.edu

Abstract: Computer science is becoming a mainstream school subject, yet we know relatively little about teaching, learning, and assessing computer science at the primary and secondary level. Few studies have followed the long-term trajectories of early computer science learners. We present a longitudinal study of a school cohort (N=48) across a three-year computer science curriculum in grades 6-8. We analyzed students' Scratch projects in terms of elaboration and computational thinking content, and modeled their association with performance on a summative open-ended assessment of computational thinking. Both metrics were associated with performance on the summative task, but engagement had a much more substantial effect. This supports the idea that early computer science experience should be designed to support students in working on personally-meaningful projects. Developing computational literacy practices may be more important for long-term growth in computational thinking than a primary emphasis on content knowledge.

Introduction
Computer science is becoming a mainstream school subject in the United States. It is being incorporated as a course in the regular curriculum (Goode & Margolis, 2011), in bite-size chunks (Wilson, 2014), and through informal communities which partially overlap with schools. For example, Scratch 2.0 topped 350,000 monthly active users for several months in 2017. The interest in computer science is spurred by a recognition of the economic opportunities afforded by computing careers as well as our societal reliance on computational media for commerce, news, work, and everyday social life. While pioneers in educational technology have argued for computers in schools for decades, the last few years have seen several well-funded, broad-reaching initiatives.

The uptake has outpaced the research. We still know relatively little about teaching and learning computer science at the primary and secondary level (Guzdial, 2008; Grover & Pea, 2013; Blikstein, 2018). Two National Research Council reports (2010, 2011) sought to define the core practices of computer science using the term computational thinking but surfaced substantial disagreement on how broad to make the definition and how it should be measured. Computational thinking certainly includes practices such as framing problems in terms of computational models and working with algorithms, but should it also include literacy practices enacted through computational media? These might include self-expression, collaboration, and developing a critical consciousness. The K-12 CS Standards (2016) adopt a compromise position by naming seven core computer science practices and designating four of them as computational thinking, "the heart of the computer science practices" (2016, p. 67). The definition of computational thinking is arbitrary but it will have real impacts on how computer science courses are assessed and taught. To find the most useful definition of computational thinking, it would be helpful to have longitudinal accounts of how students learn computer science over time. However, very little such research exists.

This paper presents a three-year longitudinal study exploring the extent to which middle-school students' programming in Scratch predicts later computational thinking skill. We developed two metrics to assess students' projects: elaboration measures how much students built out their projects, and computational content measures how much students used core computer science ideas such as control flow, events, and modularization. These metrics correspond to two different views on the nature of computational thinking: that it is broadly inclusive of literacy practices, or that it is a narrower collection of disciplinary skills.

Background
In this study, we model long-term computer science learning as growth in computational thinking, a construct whose definition is under active debate. One goal of the study is to compare alternative definitions of computational thinking. One tradition following Papert (1980) sees computational thinking as the recruitment of computers for problem-solving, but also for learning more broadly, including the learner's relationship to knowledge and sociocultural factors. diSessa (2001) gives a three-part definition of computational literacy: knowing how to interact with computational media (the material), the cognitive abilities supported by
computational media (the cognitive), and the social arrangements enacted through computational media (the social). In this view, the work of a computer scientist is inseparable from how she positions herself socially, establishes her competence, and participates in communities engaged in computational thinking practices. Research on equity and inclusion in computer science (Margolis, 2003; Barron, 2004; Margolis et al, 2010; Kafai & Peppler, 2011) tends to take this broader, practice-based view of computational thinking. These practices include computational thinking-specific practices such as incremental and iterative development, testing and debugging, reusing and remixing, abstracting and modularizing (Brennan & Resnick, 2012), as well as related sociocultural practices such as collaborating, participating in an inclusive computational culture, and communicating about computing (K-12 CS Framework, 2016).

Other researchers, largely within the field of computer science, define computational thinking more narrowly, as the distinctive skills and knowledge gained through programming experience, and which comprise the disciplinary subject matter of computer science. In this view, computational thinking is primarily concerned with designing, using, and reasoning about computational models (Aho, 2011). Wing's argument that computational thinking "represents a universally applicable attitude and skill set" (2006) frames a narrow set of skills as widely applicable. In particular, the jobs-oriented argument for expanding access to computer science tends to emphasize skills valuable to employers and not practices which deepen self-knowledge, civic engagement, or critical awareness.

These epistemological stances lead to different ways of assessing computational thinking. Studies focused on engagement, belonging, and participation in computational thinking practices tend to seek evidence of learning by analyzing artifacts designed by students (Brennan & Resnick, 2012; Fields, et al, 2016), interviewing students (Barron et al, 2013), and by asking students to engage in scenarios where students design solutions (Brennan & Resnick, 2012) or fix solutions containing errors (Werner et al, 2012). Research adopting the narrower definition of computational thinking tends to use positivistic definitions of the skills and knowledge that make up computational thinking, and seek standardized assessments whose purpose is to measure learning apart from students' situated practices (Tew & Guizdal, 2011; Tew & Dorn, 2013). Denning (2017) argues for assessing students' competencies, which he defines as "ability accompanied by sensibilities." Research comparing the effectiveness of different teaching tools or approaches often relies on un-situated measures of learning to compare learning across contexts (Armoni, Meerbaum-Salant, & Ben-Ari, 2015; Weintrup, 2016).

In this study, our goal was to measure the association between students' practices over time and their computational thinking skills at the end of the three-year curriculum sequence. The summative assessment of computational thinking (described in more detail below) required students to solve a computational problem using any tools of their choice. Students were presented with multiple cases of the same problem with larger and larger datasets, so that they became intractable without implementing an algorithmic solution. The practices assessed were comfortably within all the definitions of computational thinking discussed above.

Each of the metrics by which students' Scratch projects are evaluated may be seen as grounded in one view of computational thinking. Elaboration measures how much detail students add to their projects, regardless of whether it has anything to do with computer science concepts. Students who create elaborate projects are likely imbuing them with personal significance and taking them up in their broader social practices. On the other hand, the narrower view of computational thinking would view much of projects' elaboration as irrelevant. Computational content measures the density of blocks that map directly to core computer science concepts. For example, when a student's project contains a higher density of function definitions and function invocations, it is reasonable to assume she is exploring modularity. By comparing the association of each metric with students' later performance on the summative computational thinking task, this study explores which practices contribute to long-term growth in computational thinking.

Methods
The participants in this study were a cohort of students at an independent all-girls' middle school in the western United States, where computer science is a core required class for all three years. The teaching philosophy of the school and the computer science classes is constructionist (Papert, 1981). During much of their school day, students work on personally meaningful projects (Papert, 1991) using tools ranging from one-to-one laptops to the school woodshop and metal-working shop. They are accustomed to seeking help from each other, from teachers, and from other resources at school and online. While some tests and quizzes are given, most summative assessments take the form of projects, presentations, portfolios, and reflections. Teachers give narrative evaluation rather than letter grades, and student self-assessment appears on report cards alongside teachers' evaluations. This school environment makes it more likely that students' projects embody authentic practices and that their performance on the summative task reflects their full capabilities.
We collected student work and written reflections from this cohort throughout their three years of computer science and gave them a summative assessment at the end of 8th grade. Students worked primarily in Scratch in 6th and 7th grade. Each of the three projects analyzed from each student was the result of a curriculum unit spanning 4-6 weeks. At the end of each unit, students' projects were assessed on the use of computational thinking concepts such as control flow, events, using variables to process data, and decomposing problems with subroutines. Students also shared their projects in class and informally through their networks of followers on Scratch. Thus, in their projects, students had an incentive to attend to both subject-matter goals and to the enactment of literacy practices such as self-positioning, attending to audience, and taking up and interpreting socially-important narratives.

Students transitioned to working in Python in 7th and 8th grades (See Figure 1). Accompanying the change in programming interface was a change in emphasis from personal expression and narrative (and closer integration with their humanities classes) to modeling and conceptual exploration (and closer integration with mathematics and science classes). The summative assessment was distant from students' Scratch projects not just chronologically, but also in terms of the problem domains with which they were engaging. 48 of 67 students returned consent forms and were included in the study.

**Measures**

**Scratch**

We evaluated Scratch projects on two dimensions. *Elaboration* is defined the natural log of the total number of blocks in the project. *Computational content* is defined as the ratio of blocks from certain categories (data, events, control, sensing, and functions) to the total number of blocks (Brennan & Resnick, 2012). These two dimensions capture different kinds of interactions students had with their projects: sometimes students created detailed images or narratives, using many blocks in a straightforward (often sequential) manner. Other times, students were more focused on how their projects worked than on the end result, and tended to have fewer blocks but more concise and expressive code.

To analyze the projects, we wrote a Python package which fetches a representation of a project from the Scratch backend server and then maps each sprite, script, statement, and expression to a Python class instance, similar to the approach used by Fields et al (2016). Figure 2 shows elaboration and computational thinking distributions for each of the three Scratch projects analyzed. We calculated an elaboration and computational content score for each student by averaging the student's normalized score on the metric from each of the three projects. In the first two projects students worked from models and starter code provided by the teacher. Using normalized values allowed us to exclude the starter code from analysis, and to weight students' work equally across the three projects. Each of these metrics represents a hypothesis about what might lead to long-term effects: the extent of block-based programming practice, or the richness of the practice in terms of computational content.

**Figure 2.** Elaboration and computational thinking content in three Scratch projects.
The final project, Drawing, provided students with the least scaffolding and therefore it is unsurprising that it has greater variance on both metrics than those preceding it. In this project, students were asked to create any drawing of their choice in Scratch, but the focus of the unit was on abstraction and modularization, and one of the evaluation criteria was students’ use of functions and data to reuse code. The negative association between elaboration and computational content in this project corresponds with the intuition that code which makes effective use of data, control structures, and functions will be able to achieve a desired effect with fewer blocks. Projects with high computational content built up more complex drawing routines from reusable subcomponents, while some projects with low computational content also created elaborate effects, but they did so using hundreds of blocks effectively encoding point-to-point vector drawings. Figures 3 and 4 show two students’ drawings and excerpts from project code.

Figure 3. A student project with low computational content (z= -2.03) and high elaboration (z=1.49).

Figure 4. A student project with high computational content (z=1.40) and medium elaboration (z=0.34).

Figure 3 is characteristic of low computational content projects; the code excerpt shows superficial use of functions to break the program into subroutines, but the program itself is essentially a long sequence of imperative commands. At the same time, this is a self-portrait executed in code, in which the student depicts a detailed facial expression, gesture, hair, and clothing by using far more blocks than most projects. The student
may have felt high social significance in each line's placement. Figure 4, meanwhile, depicts a nondescript scene very similar to the model drawn by the teacher in introducing the project. This student appears to have put her energy into the structure of her project rather than its final project; the code excerpt shows an elegant use of nested subroutines with arguments. Her project is almost entirely composed of function definitions and invocations; accordingly, her project's computational content was among the highest in the class.

**Summative computational task**

At the end of the 8th grade year, we gave students an open-ended computational task over the course of two 45-minute class periods and analyzed the extent to which students used computational thinking to solve the task. The task presented students with a list of items for sale at a store, and asked students to spend a specified amount of money on exactly two items. There were six cases of the problem which only varied by the amount of money to spend and the length of the price list. The early cases could easily be completed by hand; the later cases had so many possibilities that they were intractable without the use of a computer. Then there were two final variations: a case in which three items should be purchased and a case in which two items must be purchased which added up to a certain price and to a certain weight. Students were provided each problem as a handout and were also given links to a starter Scratch and Python project initialized with the data. Students were encouraged to approach the problem as a fun puzzle, freely requesting help from their teachers and peers, and using any strategies of their choice. There was no grade, reward, or recognition attached to students' performance. At the end of each class session, students completed a survey in which they explained their work.

**Prior experience**

Finally, we attempted to control for prior experiences by including two additional factors in some models. At the beginning of sixth grade, students began by completing a series of puzzles called Carol the Robot. In these puzzles, students wrote instructions to guide a robot through a maze to collect beacons without hitting walls. We used students' scores on this task as a measure of prior computational thinking skill. Additionally, at the end of sixth grade, students' teachers were asked to estimate their general quantitative skills.

**Analysis**

We developed a rubric to evaluate students’ use of computational thinking on the summative task (See Table 1). The rubric focuses on the core computational thinking practices of framing the problem, designing an algorithmic approach, and using a computer to implement it. In evaluating students on the rubric, we used their final submissions and their reflections, in which they explained and justified the approach they took to the problem. Scoring students solely on the number of cases students solved would not have been an adequate measure of computational thinking because some industrious students spent the entire 90 minutes doing tedious but occasionally lucky guess-and-check, while other students developed computational solutions which failed due to bugs. The least successful students used manual guess-and-check with no indication of a systematic approach; the most successful students developed correct, generalized solutions in Scratch or Python. Of the students who used programming, most used Python; only three chose to use Scratch (two of whom were successful). We were surprised at the diversity of students’ approaches. Many used ad-hoc computational tools such as Excel to sort the numbers in a list or Word's search feature as part of an otherwise-manual strategy.

<table>
<thead>
<tr>
<th>Level 0</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worked by hand or using a calculator. No</td>
<td>Worked by hand or using a calculator. Used</td>
<td>Used an ad-hoc tool (Ex: Word, Excel) to</td>
<td>Attempted to implement an algorithm using</td>
<td>Successfully implemented a generalized</td>
<td>Successfully implemented a generalized</td>
</tr>
<tr>
<td>evidence of a computational strategy. (Ex:</td>
<td>a computational strategy. (Ex: decomposed</td>
<td>implement a computational strategy (Ex:</td>
<td>Scratch or Python, but did not solve all</td>
<td>algorithm solving a more complex case as</td>
<td>algorithm solving a more complex case as</td>
</tr>
<tr>
<td>guessed pairs of numbers over and over)</td>
<td>the problem; systematically</td>
<td>decomposed the problem; sorting;</td>
<td>the two-item cases.</td>
<td>well.</td>
<td>well.</td>
</tr>
<tr>
<td></td>
<td>tested cases)</td>
<td>searching)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Having defined metrics for students' Scratch projects and the summative measure, we used standard OLS linear regression to estimate the association between students' block-based programming practices and
their performance on the summative task. Because our dependent variable is ordinal, we assume the rubric measures a latent, continuous random variable. This analysis treats the rubric categories as evenly-spaced intervals, an assumption we are comfortable making because the score distribution approximates a normal distribution (M=2.75; SD=1.47; skew=-0.22; kurtosis=-0.81). For each of the two dimensions on which we analyzed students' Scratch projects, we used the average normalized score across the three projects.

**Results**

We found that both *elaboration* \((r=0.512)\) and *computational content* \((r=0.322)\) in block-based programming were positively correlated with higher performance on the summative assessment several years later. However, we found that project *elaboration* was a much more important metric (See Table 2). *Computational content* is also not a statistically significant predictor of summative scores when outliers are removed. The effects remained largely unchanged when controlling for prior computational thinking skill and teachers' estimates of quantitative skill. Figure 5 plots each metric against summative performance.

**Table 2: Regression effect sizes (and p-values) predicting summative performance**

<table>
<thead>
<tr>
<th>computational content</th>
<th>elaboration</th>
<th>prior ct (Carol the Robot)</th>
<th>prior quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80 (0.035)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.21 (0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.57 (0.094)</td>
<td>1.10 (0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.55 (0.11)</td>
<td>1.14 (0.001)</td>
<td>-0.14 (0.427)</td>
<td></td>
</tr>
<tr>
<td>0.01 (0.074)</td>
<td>1.01 (0.003)</td>
<td>-0.21 (0.236)</td>
<td>0.49 (0.118)</td>
</tr>
</tbody>
</table>

**Discussion**

We found that students who engage more deeply in creating Scratch projects, measured in either of two ways, are significantly more likely to perform well on a test of computational thinking skill several years later, even though all but three students chose to use tools other than Scratch in the summative assessment. Prior research has found little evidence of novices learning across programming interfaces (Armoni, Meerbaum-Salant, and Ben-Ari, 2015; Weintrop, 2016). There is also little evidence that programming experience develops generalized mental functions (Pea & Kurland, 1984), or for far transfer of thinking skills more generally (Barnett & Ceci, 2002). Our findings are an important result in an area where very little longitudinal research is available.

Comparing the two metrics of students' Scratch projects, we were surprised to find that students' project *elaboration* was much more associated with summative scores than *computational content*. This supports the idea that early computer science experience should be designed to support students in working on robust and sustained programming projects. Working on personally-meaningful projects may be more important than ensuring all students have a uniform foundation in basic computer science concepts. This is well-aligned with sociocultural research showing the importance of supporting the development of student interest and
identification with computer science, particularly for marginalized students. Students need to feel a sense of belonging to learn effectively, and computer science is pervasively stereotyped as being a subject most suitable for white, male, high-achieving students (Margolis 2003; Margolis et al, 2010) Fostering communities of creative media production can help to dismantle stereotypes and increase participation (Kafai & Peppler, 2011). The importance of these literacy practices is not reflected in several national initiatives to teach computer science. Future research on the reflections and self-assessments students submitted with their projects will allow us to corroborate our interpretation of the elaboration metric and to study students' process as well as product.

Our findings suggest that computational literacy practices are associated with long-term growth in computational thinking, even under the narrower definition of computational thinking. While the definition of computational thinking is arbitrary, our study suggests that students' sociocultural practices may play an important role in learning even core computer science content. Therefore, a broader definition, which recognizes computational thinking as a situated set of practices, might be the most useful. This is in line with Kafai & Burke's (2013) call to reframe computational thinking as computational participation, and diSessa's (2017) argument for extending computational thinking to computational literacy.

Conclusions
This study offers one of the only longitudinal accounts of early computer science learning, and provides strong evidence that students' programming practices lead to long-term growth in computational thinking. The breadth of our central construct, computational thinking, is still taking shape. This study also offers an empirical exploration of the effects of operating under different definitions of computational thinking. In subsequent research, we intend to strengthen these claims by analyzing this cohort's reflective writing and self-assessment over the course of their middle-school experience learning computer science.

References


Kafai, Y. B., & Burke, Q. (2013, March). The social turn in K-12 programming: moving from computational thinking to computational participation. In Proceeding of the 44th ACM technical symposium on computer science education (pp. 603-608). ACM.


Acknowledgements

Funding for this study was provided by the Lemann Center for Entrepreneurship and Educational Innovation in Brazil. We are grateful to the students and teachers who shared their time with us to make this study possible.
Collaborative Gesture as a Case of Distributed Mathematical Cognition

Candace Walkington, Geoffrey Chelule, and Dawn Woods, cwalkington@smu.edu, nchelule@smu.edu, dwoods@smu.edu
Southern Methodist University

Mitchell J. Nathan, University of Wisconsin-Madison, mnathan@wisc.edu

Abstract: Gestures have been shown to play a key role in mathematical reasoning and be an indicator that mathematical reasoning is embodied – inexorably linked to action, perception, and the physical body. Theories of extended cognition accentuate looking beyond the body and mind of an individual, thus here we examine how gestural embodied actions become distributed over multiple learners confronting mathematical tasks. We identify several ways in which gesture can be used collaboratively and explore patterns in how collaborative gestures seem to arise in a learning environment involving a motion capture game for geometry. Learners use collaborative gestures to extend mathematical ideas over multiple bodies as they explore, refine, and extend each other’s reasoning.

Introduction

Gestures – movements that accompany speech – have been found to be a powerful component of reasoning in a variety of domains (Alibali, Spencer, Knox, & Kita, 2011; Beilock & Goldin-Meadow, 2010; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004), including mathematics. The gestures learners formulate can reveal information not expressed in speech (Church & Goldin-Meadow, 1986), can show an association with conceptual performance (Goldin-Meadow, 2005; Cook & Goldin-Meadow, 2006), can be manipulated to give students new actionable ideas (Goldin-Meadow, Cook, & Mitchell, 2009; Nathan et al., 2014; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014), and, when prevented, can impaire reasoning (Hostetter, Alibali, & Kita, 2007; Nathan & Martinez, 2015). In addition, teachers use gestures to communicate ideas to students in a multimodal manner (Alibali & Nathan, 2012; Valenzeno, Alibali, & Klatzky, 2003), which may be particularly important in mathematics classrooms where gestures can make spatial and relational aspects of mathematical concepts come alive (Nathan et al., 2011).

While considerable research has been conducted on how teachers and learners use gestures during mathematical reasoning (see Alibali & Nathan, 2012 for a review), less work has looked at gesturing as a collaborative activity that is part of mathematical discussion and argumentation between different learners. Here we define collaborative gestures as gestures that physically encompass multiple learners. We propose that learners engaging in mathematical reasoning can use their bodies, particularly their hands, in collaborative ways to reinforce, extend, and redirect the mathematical ideas of others, which were also expressed through physical movement. Theories of embodied cognition (e.g., Wilson, 2002) posit that learners process and understand ideas through their bodies and their senses, and that the mind and body, rather than being separate entities, have a bi-directional relationship. A complementary theory, extended cognition (Clark & Chalmers, 1998), accentuates the idea that a student’s cognitive system extends beyond their own minds and bodies, into their environment and those around them. Collaborative gestures are a fascinating area of study for the learning sciences, because they show how cognitive processes become extended over the embodied experiences of multiple learners. In addition, collaborative gestures provide a window into understanding how the body can be leveraged to help students understand mathematical ideas as they work jointly on challenging tasks. In the present paper, we examine collaborative gestural activity that groups engage in while proving geometry conjectures.

Theoretical framework

Gesture as Simulation Action

The theory of Gesture as Simulated Action (GSA; Hostetter & Alibali, 2008) provides an empirical, embodied cognition account of how the multimodal production of gestures comes about. Gestures in the GSA framework arise during speaking when pre-motor activation, formed in response to motor or perceptual imagery, is activated beyond a speaker’s current gesture threshold. The threshold is the level of motor activation needed for a simulation to be expressed in overt gesture; this threshold can vary depending on factors such as the current task demands (e.g., strength of motor activation when processing spatial imagery), individual differences (e.g.,...
level of spatial skills), and situational considerations (e.g., social contexts). Nathan (2017) proposed an extension of GSA to account for the influences of motor activity the learner is induced to perform (e.g., directed actions) on cognition. In this reciprocating model, learners’ actions and movements serve as inputs capable of driving the cognition-action system toward associated cognitive states through a bi-directional process. In other words, in addition to cognitive states giving rise to actions, directing learners to engage in physical motions may give them new ideas and insights relevant to understanding and solving tasks. While Nathan’s (2017) extension to GSA accentuates how directing the learner to motion in particular ways can trigger new cognitive and embodied states, here we emphasize how observing and embodying the actions and gestures of others, through collaborative activity and joint sense-making, can trigger new cognitive and gestural states in learners. We take this to be an important case of distributed cognition, which we discuss next.

**Distributed cognition and collaborative gesture**

Professional practice involves the coordination of many different inscriptions and representational technologies by differently-positioned actors whose actions occur across a range of social and physical spaces (Goodwin, 1995; Hutchins, 1995). Through joint, coordinated activity, cognition becomes distributed over a patchwork of discontinuous spaces and representational media. Lave (1988) describes how “‘Cognition’ observed in everyday practice is distributed—stretched over, not divided among—mind, body, activity and culturally organized settings (which include other actors)” (p. 1). Theories of extended cognition further argue that the social and physical environment of learners is actually constituent of their cognitive system (Clark & Chalmers, 1998). The implication is that cognition, rather than existing in the head of an individual, is distributed over the bodies of multiple learners and the environment around them as they interact. One way cognition can be extended across learners is through the use of gestures that extend over multiple persons.

Prior research on learning origami has identified collaborative gestures as gestures through which a learner interacts with the gestures of a communicative partner (Funiyama, 2000). In the context of this past research, these gestures often involved a student pointing to or manipulating a teacher’s gestures about origami folds. Here we reimagine the idea of collaborative gestures to be relevant to learner-learner interactions around mathematical sense-making and take such gestures to be a case of extended cognition. We next turn to a discussion of how gestures and actions have been studied in the context of mathematical reasoning.

**Gesture, embodiment, and mathematical proof**

One type of gesture that has been identified as important to mathematical reasoning is dynamic gestures (Göksun et al., 2013; Uttal et al., 2012). These are gestures where learners use their bodies to physically formulate and then transform or manipulate mathematical objects. For example, a learner might make a triangle with their thumbs and forefingers, and then make that triangle grow, shrink, rotate, flip, etc. These kinds of gestures might be particularly important and revealing when learners are engaging in mathematical proving or “the process employed by an individual to remove or create doubts about the truth of an observation” (Harel & Sowder, 1998, p. 241). When learners successfully justify mathematical arguments, their reasoning can be transformational in nature, where they perform valid mathematical operations on objects in order to build a logical deductive chain of reasoning that transcends particular cases and applies to all mathematical objects under consideration. This kind of transformational proof scheme (Harel & Sowder, 1998) may be closely coupled with both mentally simulated and physically enacted gesture and action (Nathan et al., 2014).

Research has shown that dynamic gestures arise during the mathematical reasoning of experts (Marghetis, Edwards, & Núñez, 2014) and that dynamic gestures have a strong association with students formulating valid proofs to geometrical conjectures (Nathan et al., 2014; Nathan & Walkington, 2017). The professional practice of proof itself has been described as “a richly embodied practice that involves inscribing and manipulating notations, interacting with those notations through speech and gesture, and using the body to enact the meanings of mathematical ideas” (Marghetis, Edwards, & Núñez, 2014, p. 243). The multimodal nature of proof is also evident for novice students in classrooms, as proofs often take on verbal and gestural forms, as opposed to formal, written ones (Healy & Hoyles, 2000), and teachers and students use gestures to track the development of ideas when exploring conjectures (Nathan et al., 2011; Nathan & Walkington, 2017).

**Research questions**

Prior work has identified how gestures arise, why gestures are important, the close connection between embodied action and mathematical proof, and the centrality of gestures that show dynamic transformations. Here we extend this work by looking at how these embodied actions are used as part of a distributed cognitive system through collaborative activity across multiple learners. We address the following research questions:

1) How are gestures used collaboratively during geometric proof activities?
2) How often were collaborative gestures used, and how did this vary by expertise and type of gesture?
3) How were learners engaging differently with collaborative gestures across individuals and groups?

Methods

Participants
Participants included 20 pre-service and 34 in-service teachers, enrolled in one of five different courses at a private university. They were either enrolled in an undergraduate elementary math methods course for pre-service teachers, a graduate elementary math methods course for pre-service teachers, a graduate course for in-service middle school math teachers in their first year, a graduate class for high school math teachers in their first year, or a graduate master math teacher course for in-service teachers who are generalists interested in mathematics or who are secondary mathematics teachers. A total of 64 students across the five classes were recruited to participate, however 3 were absent on the day of the study and 7 did not consent to participate, for the final sample of 54. Forty-six participants were female, while 8 were male; 38 identified as Caucasian, 6 as African-American, 5 as Asian, 3 as Hispanic/Latin@, and 2 as Other race/ethnicity.

Procedure
Participants were placed in groups of 3 to 6 to play a Kinect-based video game for learning geometry, called The Hidden Village (Nathan & Walkington, 2017). The game included 8 tasks where players perform directed motions with their arms and then prove or disprove related geometric conjectures (Table 1). Participants were directed to take turns of the person controlling the game for each task such that everyone in the group controlled the game at least once. Due to technical issues, some groups only experienced six of the eight game conjectures. Participants were instructed to not use pencil or paper to assist them in proving the conjectures, and were told to work together. Although the body movements they had been directed to perform in the game were intended to give them insights about the proofs (e.g., for a conjecture about similar triangles they were directed to make growing triangles with their arms), the directed movements were somewhat rarely explicitly re-enacted in participants’ discussions. Here our focus is on the gestures that these learners formulated themselves to help them understand, reason through, and collaborate around the proof to each conjecture.

Table 1: Conjectures participant groups proved

| 1. The sum of the lengths of any two sides of a triangle must be greater than the length of the remaining side. |
| 2. Given that you know the measure of all 3 angles of a triangle, there is only one unique triangle that can be formed with these 3 angle measurements. (False) |
| 3. The area of a parallelogram is the same as the area of a rectangle with the same base and height. |
| 4. Diagonals of a rectangle are always congruent. |
| 5. If one angle of a triangle is larger than a second angle, than the side opposite the first angle is longer than the side opposite the second angle. |
| 6. The measure of the central angle of a circle is twice the measure of any inscribed angle intersecting the same two endpoints on the circumference of the circle. |
| 7. Reflecting a point over the x-axis is the same as rotating the point 90 degrees about the origin. (False) |
| 8. If you double the length and width of a rectangle, the area is exactly doubled. (False) |

Analysis
Video was captured of groups playing the game, and was transcribed in the Transana software (Woods & Fassnacht, 2012). Videos were clipped such that one clip was one group proving one conjecture. Each clip was coded for each gesture sequence that arose – a gesture sequence was defined as all of the hand gestures made by a single person from the time their hands rose until the time their hands returned to rest. Gesture sequences were coded as being individual if the gesturer was making a gesture that was not triggered by or related to the gestures of others, and collaborative if they were. For individual gestures, gesturers could still be building off of the speech of other learners – gesture were only coded as collaborative if learners were gesturing in response to the gestures of other learners. Collaborative gesture sequences were coded for both the learner that was performing them and the learner the performer was responding to or collaborating with. Collaborative gestures were then separated into different categories using a grounded, bottom-up approach of constant comparisons (Glaser & Strauss, 1976). As new collaborative gesture categories emerged, prior clips were revisited and recoded. Two coders completed all gesture sequence coding. Fort-four percent of the corpus was double-coded by both coders, with discrepancies, issues, and ideas discussed as coding was compared and new categories were formulated. A total of 87 clips of 12 groups across the five classes were coded for gesture sequences.
Results and discussion

RQ1: In what ways were gestures used collaboratively?

Results revealed five categories of collaborative gestures (Table 2). In the example of mirroring gestures in Figure 1, two learners are making the same triangle gesture at the same time when proving Conjecture 5 in Table 1. In the example of echoing gestures in Figure 1, one learner scales her hands in and out to make similar triangles for Conjecture 2 in Table 1, and then another learner takes up her explanation while repeating her gesture. These gestures reveal how learners can reflect on and adopt each other’s mathematical reasoning.

Table 2: Categories of collaborative gestures

<table>
<thead>
<tr>
<th>Gesture Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mirror</td>
<td>Learners make same gesture at same time, while purposefully following each other’s movement.</td>
</tr>
<tr>
<td>Echo</td>
<td>One learner purposefully repeats the gesture made by another learner.</td>
</tr>
<tr>
<td>Echo &amp; Build</td>
<td>One learner repeats all or part of another learners’ gesture, but then modifies it or builds on it in some way to show a progression of reasoning.</td>
</tr>
<tr>
<td>Alternate</td>
<td>One learner gestures their understanding, another learner responds with a different gesture showing their understanding, either building or redirecting from the original gesture.</td>
</tr>
<tr>
<td>Joint</td>
<td>Two or more learners make a single gestural representation together.</td>
</tr>
</tbody>
</table>

Figure 1. Mirroring (left) and echoing (right) gestures.

In the extended example in Figure 2, participants are discussing their reasoning for Conjecture 4 in Table 1. Tanya and Karen initially represent their understanding, with Tanya gesturing horizontal sides with her hands and Karen gesturing crossed diagonals. In their speech, Tanya focuses on the two sides being equal (Line 1), while Karen focuses on the diagonal’s opposing relationship to each other (Line 2). Lisa performs an alternating gesture to redirect the conversation to another idea – that the sides are parallel (Line 3). She then integrates Tanya’s equality gesture into her explanation by echoing it and saying “So they’re the same” (Line 5). Karen then echoes both Tanya’s equality gesture and Lisa’s parallel gesture, and builds by adding on her crossed arms gesture at the end of the sequence (Lines 6 and 8). Lisa assists her in formulating this explanation by repeating her own parallel line gesture, then echoing Karen’s crossed arm gesture (Line 7). Lisa then sweeps her diagonal hands in and out to accentuate the idea of congruency, building upon Karen’s gesture (Line 7). This transcript demonstrates how different aspects of understanding the same problem become embodied through different gestures, and how learners take up and build upon each other’s gestures to extend their understanding. Karen is able to adopt the gestures of Lisa and Tanya in order to create a mathematical argument that brings in all three of their ideas – parallel sides, congruent sides, and diagonals.

The second extended example in Figure 4 shows an example of joint gestures. Participants are discussing Conjecture 6 in Table 2, “The measure of the central angle of a circle is twice the measure of any inscribed angle intersecting the same two endpoints on the circumference of the circle.” Diana has been explaining her understanding using a gesture of two angles, one formed with her two index fingers (the inscribed angle) and one formed with her thumbs (the central angle – see Figure 3). Kyla has been somewhat removed from the conversation, and brings herself in by posing the question to Diana “Where is the circle?” (Line 1). Since it would be physically impossible for Diana to represent both angles and a circle with her hands alone, Diana responds “You have to visualize it” (Line 2). In the meantime, Kyla walks over to Diana and traces a circle around the outside of her angles, and then points to the vertex of the inscribed angle, which she believes is the outside of the circle. A third group member, Sophie, responds to the need to represent the circle in a more permanent manner by offering her hands to represent the circle, placing them under Diana’s hands (Line 3). Kyla and the fourth group member, Carl, then proceed to discuss and point to various parts of the embodied diagram that Sophie and Diana have formed (Lines 5-10). This transcript demonstrates how gestural activity can become distributed over multiple learners as they represent a single, shared mathematical system.
that is jointly embodied and that they can collaboratively refer to and discuss. The cognitive work of the mathematical system is distributed over all four of their bodies, and the gestures and speech they engage in.

1 Tanya (at the same time): I think so because two sides have to be. The two sides are equal.

2 Karen (at the same time): Yeah if they were the opposite.

3 Lisa: The two sides are parallel.

4 Tanya: Parallel.

5 Lisa: So they’re the same.

6 Karen: Wait, the two sides are parallel the height and the width.

7 Lisa: So diagonally they’d have to be the same.

8 Karen: The height and width are parallel so the diagonals have to be the same.

Figure 2. Alternating and Echo Build Gestures.

Overall, we found a variety of ways in which gestures could be used collaboratively in the sense that gestures became distributed over multiple learners. This was an initially surprising and unexpected phenomenon to arise as these students engaged in mathematical reasoning. Mirroring and echoing gestures showed one learner taking up the mathematical reasoning of another, while echo & build gestures and alternate gestures showed learners extending and refuting each other’s reasoning using embodied action. Joint gestures allowed multiple learners to use their bodies in concert to build and transform an embodied mathematical system together. These gestures did not simply accompany their mathematical arguments or their collaborative activities – they were an inexorable part of their reasoning processes.

Figure 3. (Left) Inscribed and central angles of a circle (Right) Diana’s initial gesture showing these angles.

RQ2: How often were collaborative gestures used?
A total of 443 gesture sequences were coded across the corpus, of which 218 (49.2%) were collaborative. This came to approximately 2.6 collaborative gestures per student group proof attempt. However, the number of collaborative gestures used by different groups varied widely – from 0.67 collaborative gestures per proof for one group, to 5.9 collaborative gestures per proof for another group. The use of collaborative gestures seemed to vary by teacher expertise level – teachers from the two pre-service classes used an average of 1.5 collaborative
gestures per conjecture, while teachers from the two first-year in-service classes used an average of 3.5, and teachers from the master teacher class used an average of 2.8.

1 Kyla: So where's your circle?
2 Diana: You have to visualize it.
   ((Diana making double angle gesture with thumbs and index fingers, Kyla comes over and traces a circle around it))
3 Kyla: So this is the point on your circle. Right there?
   ((Kyla points to the tip of Diana's top angle, Sophie comes over and makes a circle underneath Diana's angles with cupped hands))
4 Sophie: Yeah this-
5 Kyla: This is the outside of the circle, this is the middle of the circle. So this is where the diameter goes, right between these points.
   ((Kyla points to the top angle vertex and lower angle vertex formed by Diana, and then traces the diameter across Sophie's circle))
6 Cai: Right, right.
7 Diana: Right.
8 Kyla: So what angle are we talking about?
9 Cai: This one and these two-
   ((Cai points to three different positions on the geometric object))
10 Kyla: Are the same-
   ((Kyla points to each point on the object after Cai does))
11 Sophie: Are the same, right?

Figure 4. Joint Gestures.

Of the 218 collaborative gestures, 110 (50.5%) were echoing gestures, 56 (25.7%) were alternating gestures, 28 (12.8%) were mirroring gestures, and 27 (12.4%) were joint gestures. Of the 110 echoing gestures, 81 (73.6%) were simple echoes while 29 (26.4%) were echo and build gestures. In terms of the types of collaborative gestures the three categories of teachers tended to make, for all three teacher groups, echoing gestures made up approximately 50% of the group’s collaborative gestures (pre-service: 54.7%; novice in-service: 47.3%; master: 50.7%). Master teachers were most likely to use mirroring gestures (pre-service: 11.9%; novice in-service: 9.8%; master: 17.9%), while novice first year teachers were most likely to use alternating and joint gestures (alternating: pre-service: 23.8%; novice in-service: 28.6%; master: 20.9%; joint: pre-service: 9.5%; novice in-service: 14.3%; master: 10.4%). Overall there did not seem to be compelling differences between categories in the types of collaborative gestures made – although the sample size is small.

RQ3: How were learners engaging differently with collaborative gestures?

Figure 5 shows how often students initiated versus received collaborative gestures. Each gray box represents one group, and each pill-shape inside of a gray box represents one student. The left side of the pill shows how many collaborative gestures per conjecture the student performed (e.g., when they echoed or mirrored another learner). The right side of the pill shows how many collaborative gestures per conjecture were directed at the student (e.g., someone echoing or mirroring them). Red indicates 0 collaborative gestures by the learner, yellow indicates less than 1 but more than 0 collaborative gestures given/received per conjecture by the learner, green indicates greater than one but less than 2 collaborative gestures given/received per conjecture. The top row of the figure is the pre-service classes, the middle row is the in-service first year classes, and the bottom row is the master teacher class.
As can be seen from Figure 5, across levels of expertise, there were yellow and red students who stayed on the fringes of collaborative gesture activity. There were groups of highly collaborative green students as well. Some students tended to do a lot of individual gestures, but had group members who would echo or mirror or alternate with them regularly. This, however, was somewhat rare – most students tended to give and receive a similar number of collaborative gestures. In other cases, group members would give quite a few collaborative gestures that would get little attention from their group in terms of being considered or built upon. This figure demonstrates how cognition was embodied and distributed across student groups, and how certain learners seemed to be more central to these processes than others.

Significance
While the importance of gestures to student learning has been established in a variety of studies, less work has been done detailing how gestures allow for cognition to be physically distributed over multiple learners. Documenting this process and how it comes about with different groups of learners in different learning environments is a first step in understanding how learners can engage in joint, embodied reasoning to solve and learn from complex tasks. A variety of questions were raised by this study, some of which we’ve addressed but were beyond the scope of this paper, while others are ripe for further investigation. These include: (1) How are collaborative gestures paired with speech processes implicating different argumentation moves? (2) How are collaborative gestures associated with learners successfully overcoming trouble spots and formulating valid mathematical arguments? Are certain types or sequences of collaborative gesture more effective than others? (3) What are the tradeoffs of using gestures as a collaboration tool rather than, for instance, a shared written or digital workspace? and, (4) Can collaborative gestures be directed or taught, particularly using motion capture technologies that can detect multiple bodies in motion? This study was exploratory and thus raises more questions than it answers, but here we identify a phenomenon that seems both powerful and important to understanding and intervening upon students’ mathematical learning. This phenomenon has the potential to be leveraged in digital learning environments as motion-based technologies become more feasible in classrooms.

References


**Acknowledgments**

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A160020 to University of Wisconsin – Madison. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.
Professional Development for Secondary Science Teachers: A Faded Scaffolding Approach to Preparing Teachers to Integrate Computing

Colby Tofel-Grehl, Kristin Searle, and David Feldon
colby.tg@usu.edu, kristin.searle@usu.edu, david.feldon@usu.edu
Utah State University

Abstract: Though we often think of computing as a discrete discipline, computing competencies are integral to professional practices in STEM fields. The Next Generation Science Standards demonstrate the importance of these disciplinary tools by directly incorporating science, mathematics, and engineering with computing in grades K-12. To assist children with learning the knowledge and skills authentic to these disciplines, we need teachers with appropriate content and pedagogical knowledge of computing. However, most K-12 teachers have little to no computing background. This design-based research study explores how designing professional development (PD) with faded scaffolds and worked examples can further teacher training efforts to bring computing to core content classes. Forty-six teachers participated in summer PD opportunities that introduced them to electronic textiles as a vehicle for integrating computing into their science classrooms. The professional and curriculum development processes evolved through an iterative design intended to build better scaffolds for teacher and student learning.

As science, technology, engineering, and mathematics (STEM) education focuses increasingly on engaging students with content in ways that align with the real-world practices of the disciplines, it is clear that computational thinking and computing play major roles. For instance, the Next Generation Science Standards (NGSS) used in K-12 education in the United States identify computational thinking as one of eight key science and engineering practices. Further, evidence suggests that supporting students’ development of disciplinary identities across contexts is a promising approach to broadening participation in STEM fields (Allen & Eisenhart, 2017; VanHorne & Bell, 2017). Accordingly, most research in the learning sciences has focused largely on how to design learning environments and activities to support students in developing these identities (Bell, Van Horne, & Cheng, 2017). Hands-on making activities that merge the physical and the digital have shown particular promise in terms of supporting students’ acquisition of disciplinary practices and identities by bringing together knowledge and skills from multiple arenas of youths’ lives and providing for the investigation of real-world problems (Tofel-Grehl et al., 2017).

By definition, making activities tend to be interest-driven, open-ended, and exploratory, lending themselves more frequently to out-of-school spaces or elective courses (Peppler et al., 2016; Sheridan et al., 2014). Given their promise for supporting youths’ disciplinary identities and, thus, their participation in STEM fields, we argue that making activities integrating computing need to move into the classroom in core content areas like mathematics and science. To do so, however, requires a teaching workforce prepared with sufficient content and pedagogical knowledge to integrate making activities. Such knowledge includes an understanding of the iterative design process and computing, as well as the ability to integrate them meaningfully into core K-12 content areas. Such practices are slowly appearing in preservice teacher preparation programs, but the vast majority of inservice teachers lack such training. Further, inservice teachers have limited time to engage in professional development (PD) activities given the ongoing time demands of their classrooms. As a preliminary step towards identifying and validating appropriate scaffolds to assist science teachers in developing computational content and pedagogical knowledge, this paper reports the results of a design-based research study on teacher professional development integrating electronic textiles materials and computing in secondary science classrooms. We are guided by the following research question: How can we design professional learning environments to scaffold making and computational thinking for classroom teachers with limited computing experience?

E-textiles as a making medium for learning and engagement

Social perceptions that STEM professions are mostly the domain of Caucasian males, sometimes referred to as “locked clubhouses,” often discourage girls and underrepresented minorities when they make decisions about engaging in STEM activities (Malcom & Malcolm-Piqueux, 2013; Margolis & Fisher, 2001). Numerous approaches have attempted to broaden the participation of girls and underrepresented minorities in STEM, including mentorship, the design of more appealing activities (e.g. storytelling instead of programming; designing
Incorporating e-textiles projects into core content areas like science presents several challenges. First, the open-ended nature of the design endeavor necessitates a project-based learning approach that may not be familiar or comfortable for many teachers. During preservice preparation, extensive training in such pedagogies is uncommon, and even when teachers do receive such training, their implementation is usually limited (Simmons et al., 1999). Second, the integration of science instruction with computing is often novel for teachers. Teaching integrated STEM units, teachers can create learning environments that are more aligned with conditions outside of school, breaking down the artificial boundaries of separate subjects and creating an experience that more closely mirrors professional practice in STEM disciplines—a major goal of the Next Generation Science Standards. Integrated STEM units can also help students form deeper understandings, see the “big picture,” make the curriculum more aligned to students’ interests, and increase students’ motivation in school (Berlin, 1994; Czemiak et al., 1999). However, as with project-based learning, teachers do not usually receive training in strategies for effective implementation (Mason, 1996). Third, while teachers at the high school level are usually well-versed in their science content knowledge, very few have received training in computational thinking, including specific programming skills (Epstein & Miller, 2011; Hargrave & Hsus, 2000). As such, they need additional training to support their students’ learning appropriately.

To accommodate both the extent of training needed for teachers to be adequately prepared for an implementation of a high quality e-textiles curriculum and the limited opportunities available for professional development, the design of such training requires attention to both effectiveness and efficiency. While learning outcomes for more directed and more exploratory instructional strategies are often equivalent (e.g., Chase & Klahr, 2017; Klahr & Nigam, 2004; Likourezos & Kalyuga, 2017), the attainment of those outcomes typically occurs more quickly under higher levels of instructional guidance. Thus, to accommodate the constraints on learning time for inservice teachers, we adopted training strategies that engaged higher levels of explicit...
Faded scaffolding entails providing learners with supportive task structures and materials to guide their early learning endeavors. As they begin to master elements of the task, those supports are gradually withdrawn until learners are demonstrating successful use of the target knowledge without aid. One form of faded scaffolding that has a strong track record of success is the use of faded worked examples (Atkinson, Derry, Renkl, & Wortham, 2000). In this model, a fully worked out example is provided to learners as a narrated or annotated demonstration, illustrating a successful performance. After the worked example has been studied or rehearsed, another is provided that presents a partially completed task. Building on the previous model, the learners then perform the necessary steps to complete the partial task. In successive examples, learners take on increasing levels of autonomy by completing greater proportions of the constituent task elements without direct guidance, until they are ultimately performing effectively without direct instructional support.

The rationale for this approach is that learners can focus narrowly on specific elements of a task to master the necessary application of knowledge and skills in manageable pieces while maintaining an authentic whole-task context. By progressively expanding the scope of application for aspects of the larger task as they become more proficient, learners will neither be overwhelmed by the magnitude of a complex task, nor will they risk floundering unproductively (Renkl & Atkinson, 2003). Given the inherent complexity of managing both new content and new pedagogical strategies (Feldon, 2007), we designed professional development materials according to this model, asking teachers to begin working through fully developed curricular materials and gradually take on greater independence and creativity in crafting with e-textiles and coding the microprocessors.

Methods
The iterative development of PD detailed in this study had two primary goals: (1) the design of a professional learning environment to scaffold making skills and strategies and (2) the development of teachers’ pedagogical skills such that they could integrate computational thinking in their classrooms using e-textiles materials. These efforts were mindful that many inservice teachers not only have limited computing experience, but also may have relevant gaps in their core content knowledge. Using a design-based research (DBR) approach, we focus on designing and implementing professional development using faded scaffolding to prepare teachers for integrating computing and STEM content in the context of their standards-driven, public school classrooms. DBR is particularly useful for helping us to understand the underlying reasons why something is happening, the conditions under which a particular type of learning or interaction can take place, and the ways in which an individual’s mind interacts with the environment and any available tools. Most importantly, DBR sees interventions that change features of environments, activities or tools as part of the process to be studied. It is both prospective and reflective, meaning that designs are initially implemented based upon some hypothesized learning mechanism and means of supporting it through a particular design or design feature (Cobb et al., 2003). Later, as the design is implemented, new features emerge as salient and both design and implementation may be refined, resulting in iterative cycles of design, implementation, analysis, redesign, reimplementation, and analysis.

Participants
Over three of implementations of the professional development workshop, 108 teachers participated. Teachers came from 9 school districts across three states of the intermountain Western United States. For the initial (pilot) implementation of the PD workshop, eleven teachers from three districts in two states participated. In the second workshop, 35 teachers participated from across three states and represented eight districts. In the third implementation of the professional development workshop, we had 72 teachers from 17 districts participate. Of those 72 teachers, 10 were returning master teachers from the second professional development workshop. Of the remaining 62 participated for the first time. Teachers’ schools were located in rural or suburban districts. Districts were selected for participation based on population demographics; we sought to include teachers from highly rural spaces with high percentages of students on free and reduced lunch. Participants’ years of teaching experience ranged from 2 years to over 40.

Data collection
Qualitative data and analysis
Observations were conducted at each of the professional development workshops by experienced PD providers and documented via written and/or audio field notes. Workshops were also video and audio recorded; portions of those workshops were identified for transcription. Interviews were conducted with all participating teachers in the pilot workshop and 12 teachers at the first full implementation of the workshop. In full implementation, interviewees were selected based on the research team’s best guess about how they would adopt the curricular
materials from the PD in their classrooms. We sought to interview teachers whom we thought were likely to represent the full spectrum from high to low levels of implementation. Within this group, we also sought a range of teaching experiences and diversity by race, ethnicity, and gender. In addition to teacher participant interviews, one member of the research team interviewed members of the PD team to identify challenges from the provider perspective. Initial coding was completed using broad categories such as areas of confusion, areas requiring further practices, and areas missing needed supports.

A grounded theory (Lincoln & Guba, 1995) framework was used to establish trending patterns within both the observed responses to the professional development as well as teacher interviews. Differences between areas of challenges were discussed between research team members until consensus was achieved as the correct problem isolation and the best method of scaffolding that issue going forward.

Quantitative data and analysis

Prior to and directly after completing the professional development workshop, teachers completed a content knowledge test that intended to capture possible gaps in teacher knowledge about circuitry (DIRECT; Determining and Interpreting Resistive Electric Circuits Concepts Test; Engelhardt & Beichner, 2004). Teachers’ pre- and post-tests were scored based on the percentage of correct responses.

Overview of design and iterative development strategy

We began by developing an e-textile curriculum based on a somewhat established progression of e-textiles projects from the literature (Buechley & Qiu, 2013; Kafai et al., 2014). Further, because most microprocessors designed for sewing e-textiles projects, including the LilyPad Arduino used in the curriculum, run on the Arduino programming language, we assumed that teachers would learn Arduino (Buechley & Eisenberg, 2008). We then tested our suppositions about an appropriately scaffolded curriculum and professional learning experience for secondary science teachers through successive cycles of implementation, analysis, and redesign, spanning a smaller pilot and two years of full implementation. To document the design process, we provided participants in each professional development workshop with hard copies of the curriculum and asked them to take notes on what worked and what was confusing. At the end of each implementation cycle, we examined these notes alongside our own observations, field notes, and debriefing sessions and made appropriate changes to the curriculum prior to the next implementation. Taken together, these data sources helped us to understand where learning was appropriately scaffolded for teachers and where more support was needed. We reflected these changes in how the PD was implemented during the next design cycle and in the curriculum itself.

Our initial development of the curriculum and PD, informed by a panel of master teachers who reviewed and critiqued it, struggled with two challenges. Firstly, all parties were concerned with the length of time sewable projects would take. For both in class instruction and the professional development workshop, concerns revolved around how to provide teachers and students enough content knowledge in addition to basic training in computational concepts and practices, as well as crafting time, while still being mindful of the constraints of the school day. Master panel teachers reported that they only allocate 1-2.5 weeks to teach electricity and energy.

In addition, having teachers take a five-day professional development workshop caused concern. Because our target PD population was middle school and high school science teachers, we decided to streamline the science content sections of the PD and focus more of our time on coding and crafting with the teachers. Our pilot PD workshop ran 24 total hours with approximately 3 hours on content, 12 hours on constructing and designing the projects, and 9 hours on coding. In the initial professional development, we began by having teachers make paper circuit greeting cards, then made simple circuit bracelets with a snap switch, had teachers “hack” their bracelets, and finally made a temperature sensing lunchbox. Of all these activities, we view the “bracelet hack” (Searle, Tofel-Grehl, & Allan, 2016) as the most crucial and it has remained a mainstay of our professional development model and classroom implementations.

The “bracelet hack” is intended to better scaffold the teachers’ ability to modify code by segregating questions about functional circuitry from questions about whether or not code was written correctly. Using alligator clips, participants attach a microprocessor to a completed, functional snap switch bracelet project. By ensuring functional circuitry, the hack allows teachers to engage with code earlier in the instructional sequence without imposing a need to split attention between coding and crafting (i.e., avoiding the need to determine whether a malfunction was the result of a short circuit or a coding error).

Findings: data-driven, iterative PD development

After completing the pilot workshop, our team met to discuss the challenges and mechanisms we could engage to improve our model (see Table 1 for overview). We identified two areas of challenge. Firstly, we noted both from our conversations with teachers and from their pre-test scores that teachers lacked sufficient content knowledge
to engage the projects most successfully. Several teachers indicated during their conversations with PD providers during the construction time of the workshop that they did not feel like they knew the content on the pre-test; as one teacher stated, “I think I got a zero. Please don’t judge me for that.” Another commented, “I have a physics degree, and I couldn’t answer those questions. I wonder how my kids will do.” Specifically, we found teachers lacked knowledge of circuit types, as well as an understanding of electron transfer and how batteries store energy.

To address these issues we added time to subsequent implementations of the PD workshop to bolster the time focused on scientific content knowledge. We also moved away from the aesthetically pleasing paper circuit project of making a greeting card to having ready-made templates of the various types of circuits to facilitate teachers’ rapid prototyping of various circuits and learning about resistance and polarity more concretely. We also increased the amount of time spent on content from 3 hours to 4.5 hours.

Table 1: Summary of data-driven iterative PD development

<table>
<thead>
<tr>
<th>PD set up</th>
<th>Challenges</th>
<th>Modifications to Address Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pilot Year:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Teacher Participated in district PD on e-textiles</td>
<td>Design process was personalized but time consuming.</td>
<td>Used the deep work with teacher 1 as basis for training PD design.</td>
</tr>
<tr>
<td><strong>Year 1 PD 1:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 teachers</td>
<td>Content knowledge weakness</td>
<td>Added Content knowledge training to PD</td>
</tr>
<tr>
<td>3.5 day PD</td>
<td>Failure to engage follow-up help</td>
<td>Commenting code worksheets</td>
</tr>
<tr>
<td>Follow up support</td>
<td>Difficulty in beginning to read code</td>
<td>Introduced Coding Sandwich</td>
</tr>
<tr>
<td><strong>Year 1 PD 2:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46 teachers</td>
<td>Continued content knowledge weakness</td>
<td>Added more scaffolded classroom instruction for content knowledge-demos</td>
</tr>
<tr>
<td>4 day PD</td>
<td>Teacher</td>
<td></td>
</tr>
<tr>
<td>Follow up support</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The second area of challenge we identified from the pilot PD workshop was that teachers struggled to read even the most basic of code. Despite focused instruction from a computer scientist with extensive undergraduate teaching experience, there appeared to be a literacy gap between reading the code and understanding what it meant. While teachers could read the words, within the context of the code they had little idea what the syntax and words meant. Teachers reported feeling overwhelmed by simply looking at the code. As one teacher noted, “this looks like a foreign language, except a foreign language at least uses the same punctuation and sentence structure.” Teachers struggled with reading more than a few words of code before reporting and being observed to be overloaded (e.g., putting head on table or closing the code file). We realized we needed to begin at a much more basic place than the typical undergraduate introduction to computing; we hypothesized that because our teacher population was older than the digital natives that populate undergraduate classes, we needed to introduce teachers more slowly to looking at and reading code. Thus, we developed a system of code reading and commenting that allowed teachers to engage the code more slowly through the use of faded scaffolds.

To do this in the second implementation (Year One PD in Table 1), we used the Arduino Blink code first as a group and discussed it together, much in the way a teacher might diagram a sentence in a grammar lesson. Then, we talked with teachers about the existing comments and what they might mean. In pairs, teachers rewrote the comments in their own words. Finally, they worked in pairs to comment new lines of code based on the understandings they had developed in their earlier practice.

We also saw teachers struggle with understanding just how specific they needed to be with what they asked a computer to do. The teachers in the first professional development workshop, with the exception of one who had copious CS experience, did not understand conceptually how code works to provide instructions to the computer. To help facilitate teachers’ understanding of how code works and why it needs to be specific, we engaged teachers in the process of “coding” one of the PD staff to make a sandwich. Teachers quickly learned that when they said, “open the mayo,” the PD provider would use a knife to saw open the bottle rather than twist off the top. Teachers learned that specificity of directions is essential in computing and that each part of the code performs a very specific job.

These scaffolds added an additional two hours of support for teachers to learn coding with the hope that earlier engagement with a more scaffolded approach would facilitate better outcomes for teachers. More broadly, in teaching teachers who have never read code how to comprehend and teach the content, we have found a faded scaffolding approach that emphasizes the importance of tracing, commenting, and explaining code as a means for developing understanding (Lopez et al., 2008; Murphy et al., 2012; Teague & Lister, 2014) to be most effective.
Upon completion of the second workshop, we found that while coding outcomes were better, teachers still needed more support in understanding the physics around electricity. Therefore, we built 15-minute content acquisition podcast (CAP) videos (Kennedy et al., 2014) that we shared and discussed as part of the PD in its third iteration. We coupled these CAP videos with classroom demonstrations that allowed teachers to see activities they could use in their own instruction as worked examples. These scaffolds added an additional 2 hours of professional development time focused on content.

We also noted that teachers were still struggling with commenting and modifying code. Instead of proceeding with teaching variables and set up, we created two worksheets that provided teachers with fading worked examples of commenting and modifying code, respectively. We developed the worksheets with the intent of providing the teachers tools to document their attempts at modification and commenting. These worksheets allowed teachers to track their efforts over the course of the workshop. The modification worksheet was relatively simple. Using a T-chart design, it provided a framework for teachers to document what their coding goal was (what they wanted the lights to do) on one side and what modifications they made to the code on the other. The first line of the T-chart was filled in with an example for them; this example was discussed and attempted as a whole group during the PD to provide a fully worked example. The commenting worksheet also proved relatively simple; on the left side was the code pulled from Arduino and on the right were blank lines for teachers to fill in their own comments. Within this process, we modeled reading a line of code and writing a comment that explained what it meant. After working through a set of lines of code and commenting them as a larger group, we assigned the next ten lines of code for the teachers to comment in pairs. We checked in with the pairs and found that nearly all teachers were now capable of commenting the code with reasonable levels of accuracy. With these worksheets, the teachers ended up trying many more modifications and became more facile at knowing which pieces of code were needed to replicate for their specified goal.

After developing the professional development workshop over multiple iterations, we have found it most successful to use a three-stage faded scaffold to introduce teachers to the reading and commenting of code. In the first stage, teachers receive a piece of code for a “basic blink” program, which turns an LED on and off in one second increments, with the entirety of the comments included. The professional development leaders read and discuss the code, explicating what each line does and what the comments tell us about the code. Teachers then attempt to use the “basic blink” code with a completed project, in our case their bracelets, and modify it to make the lights blink in different sequences or frequencies. At that point, professional development leaders walk the teachers through the process of modifying the comments on the provided code so that their comments now match the modifications they have made. In the second stage, teachers receive the entire code and comments for the set up section of the code. They must then comment the lines of code that do not have comments. Answers and comments are checked for correctness and accuracy. In the final stage of the training, teachers receive a section of code and are asked to comment every line. Teacher ability and comprehension are checked a final time before teachers begin learning the next process—writing code for themselves.

The process of developing integrated scaffolds for teachers across science and computing can prove complex and arduous. However, the value of building teacher trainings and materials in this way are multifaceted. By working with teachers in this way core skills and practices across STEM disciplines are integrated in authentic ways. It also provides teachers a model of successful strategies and instructional practices they can implement in their classrooms. By engaging teachers in learning through faded scaffolds and worked examples we can influence their use of those same instructional support strategies in their classrooms.

**Discussion**

Within the context of teacher knowledge there were several cogent trends distilled from the data. First, we observed multiple knowledge gaps related to teacher content knowledge around circuits and electricity. Despite being secondary teachers holding degrees in relevant content areas, teachers often lacked the knowledge to make hands-on projects involving circuits successfully. In addition to a lack of science content knowledge, teachers also lacked computing knowledge. While this was expected, we were surprised to find that the level of content standard for introductory computer science at the undergraduate level far surpassed teachers’ comfort levels.

In thinking about efforts to support teachers with the myriad of challenges they face in classrooms, researchers and professional development providers must be mindful of several things. First, it is important to manage teacher learning in a fashion that teachers feel supported and secure in moving from the role of teacher to learner. While we may argue that the roles of teacher and learner ought be intertwined, many educators’ self-concept as teachers involve the belief that they need to be an expert and knowledge giver. For teachers, being able to correctly answer student questions is not just an issue of pride, but an issue of identity. Another issue for consideration is managing the amount of learning and load with which teachers engage during professional development. Such PD often asks teachers to learn new pedagogical skills, curricular content, and disciplinary
approaches. This quantity of information imposes a highly taxing load on teachers that can leave them frustrated and insufficiently prepared to adopt new approaches to integrated learning effectively within their classrooms. By engaging teachers in professional development that was constructed to scaffold learning using fading worked examples, teachers are provided multiple advantages in their professional learning. In addition to benefitting their own learning, those same scaffolds and examples can serve as meaningful tools for their students during classroom instruction.

By using faded scaffolds and worked examples in teacher training, we were able to provide teachers training on a complex yet manageable set of interrelated projects that incorporate STEM practices along with classroom science content. As we integrate the content and the practices for teachers in their own trainings, future work can explore the way teachers engage and model this integration process in their own instructional practices.

References


Acknowledgments
This study was made possible by a grant from the National Science Foundation (award #1542801). 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 National Science Foundation. Special thanks go to Yasmin Kafai, Kylie Peppler, Archie Holmes, Helen Hu, and Gloria Townsend for their support and mentorship in bringing the many facets of this project together.
Measuring the Scale Outcomes of Curriculum Materials

Natalie Pareja Roblin, University of Pittsburgh, natalie.pareja@gmail.com
Christian Schunn, University of Pittsburgh, schunn@pitt.edu
Susan McKenney, University of Twente, susan.mckenney@utwente.nl

Abstract: Learning sciences research often focuses on learning environments that aim for substantial impact on learners and is increasingly concerned with doing so at scale. While learning scientists regularly measure teacher outcomes and learner outcomes, data are rarely collected on other dimensions of scale that are important for uncovering practices and designs that achieve wide-reaching, long-term impact on learning, notably: spread of underlying norms and beliefs, sustainability of use, and shift of ownership into the hands of the practitioners using the learning environments. This paper describes an instrument for measuring these scale outcomes with a focus on a particular component of learning environments: curriculum materials. By presenting both the instrument and results from its pilot use, this paper offers a conceptual as well as a methodological contribution toward developing much-needed instrumentation for evaluating the outcomes of learning environments at scale.

Introduction
A core commitment of the learning sciences is to envision and create learning environments with transformative outcomes. A key challenge to upholding this commitment has been the development of environments that both substantially impact the learners and do so in ways that are sustainable at scale. To date, the learning sciences have problematized what it means for learning environments to be productive across many contexts (Tatar, Roschelle & Hegedus, 2014). Further, the field has also conceptualized the learning of the system that occurs when attending to various stakeholders in a larger ecology, such as principals, districts, states (Marx et al., 2004). Yet in order to have a science of scaling, we need to develop measures of scale that are meaningful across specific project instances. This is no easy task. By their very nature, scale outcomes are broadly distributed across place and time (i.e., potentially across the world and across years), so the feasibility of measures is particularly problematic. This paper describes the Scale Outcomes Rubric, an instrument to measure the outcomes of curriculum materials as well as initial results from pilot testing the rubric. As such, this paper offers both a lens and a tool for collecting data on scale outcomes of curriculum innovation through materials. Here, the term curriculum materials, refers to resources designed for use by teachers in the classroom to guide their instruction, including textbooks, supplementary units, and instructional media (Remillard, Harris & Agodini, 2014).

Theoretical framework: A multidimensional view on scale
Impacting learning at scale of major interest for education policy, practice, and research. This interest stems from the desire of education policymakers and curriculum developers to create deep and lasting changes in teaching and learning (Sanders, 2012; Sabelli & Harris, 2015), and is reflected both in a growing literature concerned with scaling up educational innovations (Coburn, 2003; Lee & Krajcik, 2012; Levin, 2013; Looi et al., 2014; Lynch, Pyke & Grafton, 2012; McDonald, Kessler, Kaufman, & Schneider, 2006; Sanders, 2012), and in increasing funding for research and development of innovative interventions that are brought to scale (e.g., Investing in Innovation program at the US Department of Education, ITEST SPrEaD program at the US National Science Foundation). Most research concerned with the scaling up of curriculum innovations has focused either on the provision of convincing evidence of their effectiveness in various settings and with a variety of teacher and student populations (e.g., Geier et al., 2008; Lynch et al., 2012; Plass et al., 2012; Tartar et al., 2008), or more generally on the challenges of going to scale (e.g., Datnow, 2002; Lee & Krajcik, 2012; Levin, 2013; Lynch, 2012). In these, scale has been viewed primarily as increasing the number of implementing schools and districts.

Others have argued that going to scale cannot be undertaken after proof-of-concept, but that it is an essential concern that must be factored into the initial innovation design (Clarke & Dede, 2009; McKenney, 2018). Doing so requires a more comprehensive view of scale that acknowledges its complexity and attends to the qualitative changes in educational practice resulting from the reform (Coburn, 2003; Cohen & Ball, 2006; Elmore, 1996; Sabelli & Harris, 2015). To capture this comprehensive view, Coburn (2003) proposed a multidimensional conceptualization of scale that includes attention to the depth of changes in classroom instruction; to issues of sustainability; to spread of norms, principles and beliefs; and to shift in ownership from external designers to schools and districts so that reform can become self-generative. These dimensions are taken as points of departure for creating an initial framework for measuring the outcomes of (science) curriculum materials. Learner outcomes

ICLS 2018 Proceedings 568 © ISLS
and teacher outcomes—critical to the depth of changes in classroom instruction—are frequently measured (though teacher outcomes are measured much less often). Dimensions of sustainability, spread, and shift in ownership are of critical importance for supporting learning at scale, but rarely measured. We initially operationalized these as follows:

- **Sustainability**: Curriculum functions without implementation scaffolds put into place by the designers.
- **Spread**: Curriculum or underlying ideas have been adopted for use by at least twice as many units (teachers, schools, districts) as were involved in the pilot testing.
- **Shift**: Systems are in place for local maintenance of curriculum implementation.

We then undertook the study described here to further conceptualize and measure these three dimensions.

**Study purpose**

A nuanced view of innovations at scale acknowledges the complexity of multiple dimensions involved. An important first step in deciding whether or not an innovation holds potential for implementation at scale, is already being taken through evaluation studies of curriculum materials which focus primarily on demonstrating effectiveness with respect to student learning, and (though less often) teacher learning. Yet, understanding scale also requires consideration of sustainability, spread of key pedagogical principles and beliefs, and shift in ownership (Coburn, 2003; Clarke & Dede, 2009). Tools to systematically describe and assess the outcomes of curriculum materials on these three dimensions of scale are currently lacking. At the same time, they are urgently needed. To design effectively for scale, we need feedback on the characteristics of our designs which enable or hinder innovations in reaching large numbers of learners across multiple contexts. While designers may be able to envision and create learning environments with transformative outcomes in the short term, they need to be able to test sustainability, spread and shift to understand if and how their designs thrive in the long term. To address this need, the present study set out to develop an instrument to efficiently describe the outcomes of curriculum materials on sustainability, spread, and shift in ownership. The instrument was validated with a sample of K-12 research-based curriculum materials for K-12 science developed with federal funding in the United States.

**Methods**

**Sample**

To test the instrument, a broad sample of research-based curriculum materials was obtained, focusing on those funded by the National Science Foundation (NSF) and the Institute of Education Sciences (IES). We focused on these grant-funded materials because 1) NSF and IES are the largest funders of research-based curriculum materials in the United States (Feder, Ferrini-Mundy, & Heller-Zeisler, 2011), and 2) the extra accountability associated with public funds would likely motivate projects to document evidence of impact. The sample was limited to projects awarded between 2001 and 2010 to ensure access to documentation on scale outcomes.

We obtained and analyzed in detail documentation from 51 projects concerned with the design of K-12 science curriculum materials for classroom use. The identification and selection of these projects followed a five-step procedure. In the first step, all NSF and IES grant awards concerned with curriculum design for K-12 science education were sought in the official databases of each funding agency. This resulted in 1,301 hits. In the second step, project abstracts were screened to identify relevant awards, using the following inclusion criteria: the project targets mainstream K-12 science education, has curriculum design as an important goal, and focuses on curriculum materials for classroom use. This narrowed the selection to 162, after which different awards linked to the same project were then merged in step 3, yielding 146 projects. To make the process of in-depth analysis of project scale outcomes manageable, step 4 involved a random subsample of projects based on principal investigators. To ensure efficient use of principal investigators’ time for member-check interviews, when a principal investigator received multiple awards within our 10-year time window, all the awards were included, resulting in 83 projects. We then (step 5) carefully examined available documentation (e.g., academic publications, evaluation reports, sales figures) that could be reporting on the outcomes of the curriculum materials, which reduced the sample to 51 projects with any documented curriculum outcomes (some projects focused on theory testing) and 26 projects with any evidence of sustainability, spread, and/or shift in ownership. These materials were highly diverse, ranging from elementary through high school, from short modules to multi-year sequences, from textbooks to simulations, and from primarily student focused materials to also including many forms of rich teacher support.
Instrument development

Building on our theoretical framework, we developed an instrument to describe the outcomes of science curriculum materials on three dimensions of scale typically understudied: sustainability, spread, and shift in ownership. Feasibility was an important consideration for instrument development, which followed an iterative process and was informed by 1) a literature review (only briefly summarized in the theoretical framework section), 2) a systematic analysis of documentation produced by a sample of research-based science curriculum development projects, and 3) feedback from experts in the field of (science) education research and curriculum design. First, key components were identified for each of the three dimensions of scale based on relevant literature about scaling up educational innovations. Next, inductive analyses of documentation produced by a small subset of science curriculum development projects (N=3) that were identified to have successfully achieved scale were used to refine our operationalization of each dimension and to identify the types of outcomes evidence typically reported by development projects. This resulted in a first draft instrument that was then discussed with experts in the fields of science education research and curriculum design (for more details, see validity and reliability section below). Based on the expert feedback, further refinements were made to the instrument. The final version of the instrument comprises three main sections that describe the outcomes of curriculum materials on sustainability, spread, and shift in ownership. Each dimension is operationalized into a set of two or more components that describe key project outcomes theoretically linked to the respective dimension (see Table 1).

Table 1: Operationalization of Sustainability, Spread and Shift in Ownership Outcomes in the rubric

<table>
<thead>
<tr>
<th>Components</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sustainability</strong></td>
<td></td>
</tr>
<tr>
<td><em>Intention to continue (partial)</em> use of curriculum materials</td>
<td>Individual teachers express <em>intent</em> to continue use of the curriculum materials after direct project support ends.</td>
</tr>
<tr>
<td><em>Sustained (partial) use of curriculum materials</em></td>
<td>Individual teachers <em>continue</em> use of the curriculum materials or parts of them after direct project support ends.</td>
</tr>
<tr>
<td><em>Sustained use of key pedagogical ideas</em></td>
<td>Individual teachers express <em>intention to or continue use</em> of key pedagogical ideas underlying the curriculum materials after direct project support ends.</td>
</tr>
<tr>
<td><strong>Spread</strong></td>
<td></td>
</tr>
<tr>
<td>Spread of curriculum materials</td>
<td>Curriculum materials are used by new teachers (not previously involved in the project), possibly in other grade levels than originally intended or in other subject areas than originally intended.</td>
</tr>
<tr>
<td>Spread of key pedagogical ideas</td>
<td>Key pedagogical ideas, not previously in place, are used by new teachers (not previously involved in the project) and/or in other grade levels or subject areas than originally intended.</td>
</tr>
<tr>
<td><strong>Shift in ownership</strong></td>
<td></td>
</tr>
<tr>
<td>Formal decision to adopt the materials after direct project support ends</td>
<td>Decisions made by a district, school, or department to adopt the curriculum (or key practices/features of the curriculum) beyond the pilot or field trial.</td>
</tr>
<tr>
<td>Formal decisions to adapt the curriculum materials to local needs</td>
<td>Decisions made by a district, school, or department to adapt (components of) the curriculum to local needs.</td>
</tr>
<tr>
<td>Maintenance of professional development and physical support structures</td>
<td>Local systems are put into place for continued teacher professional development (e.g., school-based professional development, district level supervisors or coaches) and/or maintenance of the physical aspects of the curriculum (e.g., kit refurbishment, website access, etc.).</td>
</tr>
<tr>
<td>Presence of local champions</td>
<td>Local champions (e.g. teachers, parent groups, change agencies, district administrators) support the dissemination, adoption and implementation of the curriculum materials.</td>
</tr>
</tbody>
</table>

Content validity refers to the extent to which a measurement instrument captures all the facets of a construct (Wilson, Pan, & Schumsky, 2012). One way of achieving content validity involves a panel of subject matter experts that independently consider the importance of individual items within an instrument (Ayre & Scally, 2014). In the present study, feedback was sought from experts in the field of (science) education research and curriculum design in order to validate the rubric. The validation procedure involved two steps. In a first step, science education researchers were asked to 1) assess the importance of the components identified to describe
Data collection and analysis

To identify and select relevant project documents, studies and reports produced by the 51 sampled projects were sought on project websites, principal investigator’s personal websites, and scientific databases (Google Scholar and Web of Science). Keywords used to search for project documentation in scientific databases included: award number, project title, and/or (co- principal investigator’s name. In a member-check interview, principal investigators were asked to go over the search results and identify additional publications that could be relevant to our analyses. This resulted in the identification of over 500 references to publications and project documents. We were unable to obtain the full text for 43 of these references, and the majority of these missing references (53%) were conference presentations.

Available publications and documents were then screened to determine relevance to our study. To be deemed relevant, documents had to report on project outcomes related to sustainability, spread, and/or shift in ownership (see Table 2). In order to ensure a comprehensive review and prevent potential publication bias, all documents reporting on the scale outcomes of the project, regardless of the publication type, were included in the analyses (e.g., journal articles, book chapters, conference proceedings, evaluation reports, sales reports, web analytics). However, when multiple document types reported results from the same study, peer-reviewed and/or most recently published sources were prioritized, and duplicates were removed.

Next, documents were screened for analyzability. Available project documentation was reviewed to describe the project outcomes on three dimensions of scale: sustainability, spread, and shift in ownership. When a document reported multiple studies, results for each study were coded separately. Document screening resulted in the identification of 42 documents reporting on sustainability, spread, and/or shift. Only 16% of the documentation reporting on these dimensions was peer-reviewed (e.g., journal articles, edited book chapters,
conference proceedings), while the largest proportion consisted of non-peer reviewed sources including project (evaluation) reports (62%), web or sales analytics (17%), or unpublished manuscripts (5%).

Finally, documents were analyzed. Because of the limited evidence available, project outcomes for each of the dimensions were scored as either absent or present, although more levels of each dimension were originally conceptualized (e.g., sustained use in only a few contexts vs. sustained use in most contexts). Table 2 contains examples of the types evidence collected for each of the dimensions studied.

Table 2: Examples of project evidence for outcomes related to sustainability, spread and shift

<table>
<thead>
<tr>
<th>Component</th>
<th>Indicator</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sustainability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention to continue (partial) use of curriculum materials</td>
<td>Individual teachers express intention to continue use of the curriculum materials after direct project support ends.</td>
<td>“In the last year of the program, the second cohort of teachers unanimously stated that they plan to continue using the materials after the program officially ends. In fact, they were very concerned about retaining access once the funding was over.”</td>
</tr>
<tr>
<td>Sustained (partial) use of curriculum materials</td>
<td>Individual teachers continue use of the curriculum materials or parts of them after direct project support ends.</td>
<td>“Even as the grant funding has decreased, teachers and students continue to use resources developed during this ITEST project. The curricular materials, as intact projects or as discrete activities, are currently in use by 8th, 9th, and 10th grade teachers and students.”</td>
</tr>
<tr>
<td>Sustained use of key pedagogical ideas</td>
<td>Individual teachers express intention to or continue use of key pedagogical ideas underlying the curriculum materials after direct project support ends.</td>
<td>No evidence available.</td>
</tr>
<tr>
<td><strong>Spread</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spread of curriculum materials</td>
<td>Curriculum materials are used by new teachers (not previously involved in the project), possibly in other grade levels or subject areas than originally intended.</td>
<td>“Due to the results of the impact of the program on student outcomes in reading on the state reading test as well as the positive findings from the teacher surveys, the district is expanding the program to all 72 buildings.”</td>
</tr>
<tr>
<td>Spread of key pedagogical ideas</td>
<td>Key pedagogical ideas, not previously in place, are used in other grade levels or subject areas than originally intended and/or by new teachers (not previously involved in the project).</td>
<td>“In terms of the effects of participation on the respondents’ teaching in other areas, the teachers answered with a significant (meaning, with a greater-than-expected consensus) that participation had a moderate effect on their teaching other content (…) When asked to elaborate on what the changes were, respondents tended to cite an increased use of inquiry, or uses of the engineering design process applied to areas not normally characterized as unique to engineering.”</td>
</tr>
<tr>
<td><strong>Shift</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal decision to adopt the materials after direct project support ends</td>
<td>Decisions made by a district, school, or department to adopt the curriculum (or key practices/features of the curriculum) beyond the pilot or field trial.</td>
<td>“The Director of Curriculum and Training and the [name of curriculum] specialist at [publisher] also reported that the materials are being distributed throughout the nation, with large school districts making adoptions of [curriculum X] (e.g., in Arizona, Oregon, Georgia, New York, Illinois, and Ohio).”</td>
</tr>
<tr>
<td>Decisions to adapt the curriculum materials to local needs</td>
<td>Decisions made by a district, school, or department to adapt (components of) the curriculum to local needs.</td>
<td>No evidence available</td>
</tr>
</tbody>
</table>

ICLS 2018 Proceedings
Maintenance of professional development and physical support structures

Local systems are put into place for continued teacher professional development and/or maintenance of the physical aspects of the curriculum (e.g., kit refurbishment, website access, etc.).

“…additional help with facilitating professional development workshops was provided by consultants from [Publisher], but the internal leadership team was the primary deliverer of the professional development. This leadership team met monthly to plan professional development for the other teachers. These planning sessions were focused on the identified needs of teachers and students in the district.”

Presence of local champions

Local champions (e.g. teachers, parent groups, change agencies, district administrators) support the dissemination, adoption and implementation of the curriculum materials.

“Texas teacher leaders have presented the results of their work in the [project name] project at geoscience professional meetings […] The teacher leaders are not only serving as effective [project name] Climate ambassadors to achieve broad dissemination of the [project name] Climate modules but also are evolving into respected colleagues, as evidenced by their professional accomplishments.”

Results

As shown in Figure 1, roughly a quarter to a third of the studied projects had any evidence on each of sustainability (33%), spread (29%), and shift (27%).

![Figure 1](ICLS 2018 Proceedings.png)

Figure 1. Proportion of projects with evidence of sustainability, spread and/or shift in ownership (N=51).

Table 3 shows the percentages of projects with any evidence on each component of sustainability, spread, or shift. Within each dimension, degree of availability varied substantially across components.

Table 3: Proportion of projects with available evidence for each component (N=51)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Components</th>
<th>Proportion of projects with available evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustainability</td>
<td>Intention to continue (partial) use of curriculum materials</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Sustained (partial) use of curriculum materials</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Sustained use of key pedagogical ideas</td>
<td>0%</td>
</tr>
<tr>
<td>Spread</td>
<td>Spread of curriculum materials</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>Spread of key pedagogical ideas</td>
<td>4%</td>
</tr>
<tr>
<td>Shift in ownership</td>
<td>Formal decision to adopt the materials after direct project support ends</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>Formal decisions to adapt the curriculum materials to local needs</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Maintenance of professional development and physical support structures</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>Presence of local champions</td>
<td>24%</td>
</tr>
</tbody>
</table>
Table 4 demonstrates how we might examine scale outcomes in light of other information. In this case, the scale outcomes are examined in relation to start year of the award and the total amount awarded. This particular 10-year funding window experienced substantial changes in funding priorities (e.g., a shift from allowing development-only to requiring research alongside development). Additionally, one might have predicted that high levels of project funding is required to aim for broad use and have resources to study scaling outcomes. Only evidence of sustainability showed any changes over time, with projects in the second half of the time window being 30% more likely to provide evidence of sustainability. Follow-up analyses showed this change was the result of shift towards materials that could be distributed online, since web-metrics provide easy access to sustainability information; interestingly, though, this did not lead to significantly greater information about spread. Finally, there was no evidence that larger grants produced more scaling outcomes evidence; the larger projects focused the project on other aspects of the work (e.g., developing more materials or conducting more research on student and teacher outcomes). Overall, these analyses show how the instrument can give insight not only into specific project outcomes, but also how it might be used to investigate the scale outcomes of funding programs and mechanisms.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Exp(B) Evidence of Sustainability</th>
<th>Exp(B) Evidence of Spread</th>
<th>Exp(B) Evidence of Shift in Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Award Start Year</td>
<td>1.30*</td>
<td>1.12</td>
<td>.92</td>
</tr>
<tr>
<td>Total Amount Awarded</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p<.05

Discussion
This study was undertaken to develop an instrument to describe the outcomes of curriculum materials on sustainability, spread and shift in ownership. The content validity results seem promising, but further research is necessary to assess and refine its reliability between multiple coders. In so doing, it would be useful to seek lessons learned during the development of related instruments. Although we know of none that specifically measure sustainability, spread and shift, the Designing for Sustained Adoption Assessment Instrument (Stanford, et al., 2016) and i3 Sustainability Rubric (https://i3community.ed.gov/insights-discoveries/1835) do seem relevant.

The general lack of evidence is a notable, though not necessarily surprising finding. In light of recent criticism about the limited evidence available about the actual scale outcomes of development projects (Stanford et al., 2016), our findings suggest that this is not only due to lack of tools for measuring such outcomes, but also because few projects pay attention to them in any formal way. In the interviews, most PIs expressed frustrations with wanting to know what happened to their work but not having any means to investigate that, often noting that the funding requests for proposals paid little attention to scale and therefore limited how much of the budget could be devoted to these aspects. Therefore, this issue seems an important consideration for funders. We take the stance that all stakeholders would benefit if funders were to support impact studies of both federally-funded and commercial curriculum materials.

Further research on scale outcomes could help both curriculum designers and policymakers alike. For designers, it would be useful to know: What characteristics of curricula contribute to positive outcomes at scale? Equally? Or are particular curriculum features more regularly associated with sustainability, spread or shift? For policymakers, broader analysis of funded curriculum design projects and their outcomes could give important feedback on past funding support and help identify key concerns to be taken up in future programs that support the design of learning environments in general and of curricula in particular.

This study takes a first step toward measuring the scale outcomes of curriculum materials. Although, the Scale Outcomes Rubric was tested with grant-funded K-12 science curriculum materials, it could be used in other areas and possibly also for higher education. The instrument may help to frame new data collection on the impact of curriculum materials, or to develop models of factors that may contribute to yielding the desired outcomes in the long term, after innovations are left to thrive on their own. While more work is needed, this study brings us one step closer to being able to measure the extent to which innovations are able to reach large numbers of learners across multiple contexts.
References
Developing Assessments That Measure Core Ideas and Scientific Practices: Challenges and Insights

Veronica L. Cavera, Ravit Golan Duncan, Clark A. Chinn, and Moraima Castro-Faix
Veronica.Cavera@gse.rutgers.edu, Ravit.Duncan@gse.rutgers.edu, Clark.Chinn@gse.rutgers.edu, Moraima.Castro@gse.rutgers.edu
Rutgers University

Abstract: Current science reforms in the US and elsewhere advocate that students should learn core disciplinary ideas through engagement in scientific practices. Engaging students in practice allows for a deeper learning of content and a more productive understanding of how scientific knowledge is developed. In order to evaluate the effectiveness of student learning in practice, assessments have to simultaneously address students’ understanding of content and practice. In order to investigate the kinds of heuristics that may assist in developing such assessments, we present data from parallel phenomenon-based modeling assessments given to 11th grade biology students. While students constructed structurally similar arguments, the amount and how content knowledge was used differed. Students were able to bring in content knowledge more readily when presented decontextualized phenomenon-based models and used this knowledge to make productive connections between evidence and provided models. We discuss heuristics and implications for design.

Background
A critical aspect of current science reform is content learning through the practices of science. Doing so gives students the opportunity to not only demonstrate what they know (the content) and how they know it (the practice) but to also do so in an authentic environment (Pellegrino, 2012). One such document is the Framework for K-12 Science Education (NRC, 2012) accompanied by the Next Generation Science Standards (NGSS Lead States, 2013). Competence within the framework comprises three major elements: (1) Disciplinary Core Ideas, the big ideas within the domain, (2) Scientific and Engineering Practices, the practices that scientists engage in such as argumentation, modeling, investigation, etc., and (3) Cross Cutting Concepts, ideas that define the nature of what it means to do and know science that crosses disciplinary boundaries (NRC, 2012). Each of these individually is insufficient to define competence, but the combination can more accurately express it through performance expectations. Performance expectations give a more specific definition of what is expected of students at specific grades and grade bands as expertise develops. These expectations provide more specificity about the practices of science used and demonstrate how these integrate with core content. Ideally, these performance expectations should lead to coherence among curriculum, instruction, and assessment design (Mislevy, et al., 2017). Assessments are designed with the purpose of assessing whether students are reaching these desired outcomes. They can provide a clear and powerful tool in improving science education. However, if assessments are not well-aligned to the content or the practices under evaluation, they can have unintended consequences (e.g. Pellegrino et al., 2001; Ruiz-Primo et al., 2012). The concept of a performance expectation does provide some guidance in what to assess particularly as science assessments are often viewed as measuring whether students know a grade-level topic (Pellegrino, 2013).

As a community, we recognize the challenges around designing assessments and have begun to develop ways that help rigorously evaluate student cognition (e.g. Mislevy, et al., 2017; NRC, 2014; Pellegrino, 2001; 2012). While there has been a push for making the process of design more rigorous, there is less work around what these designs need to look like, and how such designs should be framed. More explicitly, we need to make the heuristics for designing science assessments more salient so that as a community we can communicate what features and how these features should be made explicit to students. In our own work, we tried to assess two dimensions in the context of genetics using an extended task in which students wrote an argument wherein they evaluated two models of a phenomenon. We found that students engaged with our task in different ways depending on how the mechanism within the models were explained and provide heuristics when students engage in a modeling task.

Theoretical framework
Current science reforms task designers with creating assessments that assess students’ ability to integrate content knowledge with practice. These tasks are often given under the larger presentation of some complex phenomenon in order to make the activity authentic (NGSS Lead States, 2013; NRC, 2012). In order to contend with this,
guidelines are provided with detailed bounds of what students should know and how they should demonstrate that knowledge. In the case of the NGSS, these are provided in the form of performance expectations (Pes). PEs provide details such as the boundaries of what content students should know at each grade band, clarification about the underlying scientific practice under evaluation and how both should be evaluated in the context of the overarching disciplinary idea. Each PE is helpful but leaves several questions about the claims that can be made about what students should know and how to assess those claims. The design of valid and reliable assessments requires not just understanding the boundaries of the required practice and content under focus but going beyond that level of content for how students will engage with the materials. Pellegrino (2013) notes that beyond content, designing science assessments requires attention to aspects like how students will conceptually engage with models and evidence, and how students will engage at different grade levels and across various medium (e.g. computer-based or paper and pencil). Collectively, it is critical to consider the kinds of evidence obtained from assessments and the conclusions about student proficiency.

In order to more rigorously link the content and practice with the task, designers have turned to more systematic approaches like Evidence Centered Design (ECD) (Mislevy & Haertel, 2009). The process can be summarized into three larger components; the claim made about the knowledge expected from students and how they are expected to demonstrate it, evidence that students have demonstrated knowledge, and the tasks performed by students that show their knowledge (NRC, 2014; Pellegrino, 2013). PEs serve as claims about student proficiency while student performance on items provides evidence of student competency. This relationship needs to be unambiguous so that valid interpretations can be made about student performance.

Previous work into this kind of rigorous assessment design has been largely focused on providing a roadmap of how to go about ECD (e.g. NRC, 2014; Pellegrino, 2012, 2013; Mislevy, 2017). This work is important for exploring how to better integrate and connect these dimensions. This framework and others like it provide a critical process for design but do not help designers address the specific heuristics that underpin how students should and can understand specific aspects of the task. Increasing the rigor between each aspect of the design of assessments is critical, and cannot be understated. However, how students interpret the specific feature of science assessments, and how these features can be designed to contend with student interpretation is less well studied. As it stands, we need to provide insight into the kinds of guiding principles that help with the interplay between designing assessments that elicit evidence of content knowledge and practice by making clear guidelines about the features and aspects of tasks that students attend to during such tasks. Doing so should help not only identify how students interact with assessments, but with improving the sensitivity of assessments to engage students around a variety of topics and contexts. Providing heuristics that help bridge these sorts of gaps will help better connect what we as designers need to know for design so that we can create more efficacious assessments that better assess students understanding of science.

In our own work, we assessed students about a genetics context (Duncan, Choi, & Castro-Faix, 2016). We developed assessments that aim to evaluate students’ competencies and reasoning with models, and the construction of evidence-based arguments about core genetics topics in molecular and classical genetics. We designed counter-balanced, structurally similar assessments that had students arguing about two models of a genetics phenomenon. One assessment focused on molecular genetics ideas, and the other focused on classical and content knowledge? What are some design heuristics that can inform the design of NGSS-aligned assessments?

**Methodology**

**Study context**

The study took place in a North Eastern suburban high school with five biology teachers and their 11th grade students (n = 271). The school was relatively diverse and 34% of the students were eligible for free or reduced lunch. Instruction lasted 10 weeks and addressed topics in genetics. Of these, five weeks were dedicated to molecular-centered lessons and included topics about the central dogma of DNA and the fundamental role of proteins. The other five weeks of instruction included key concepts about classical genetics and covered topics related to patterns of inheritance. Students were given multiple opportunities to develop extended evidence-based arguments in support of a chosen model (from 2-3 alternative models). During instruction, students were provided scaffolding which afforded opportunities to build a class consensus list of what qualifies as a good model (Pluta, Chinn, & Duncan, 2011). A feature of the instruction tasked students to make connections between the evidences and models (e.g. Chinn, Reinhart, & Buckland, 2014). Students used this list in order to engage in epistemic
reasoning of what qualities are important in scientific models. These public criteria for good models was displayed for the duration of instruction. All curriculum engaged students in the development, evaluation, and revision of models of genetic phenomena.

**Instrument design and data collection**

The assessment included two parts and lasted for 60 minutes. In the first part of the assessment, students were provided three pieces of evidence and tasked with generating a model of a phenomenon. Students were then given three more pieces of evidence and two models, and tasked with building an argument in support of the model they thought was best. We will be discussing data from the second part of the assessment, the argument. For brevity, we will refer to these assessments as the Molecular Modeling Assessment (MMA) and the Classical Modeling Assessment (CMA). The evidence provided to students in each assessment were either of high quality (e.g. scientific studies), provided information on the phenomenon (e.g. a description of the symptoms) or was an anecdotal account. Figure 1 provides a sample evidence from each assessment. Evidence 1 (left) from the MMA describes that a mutated gene causes DEB (the disease under study) and this mutation may change the protein. Evidence 3 (right) describes the ratio of healthy, to individuals with the disease, to those who died from FHD (the disease in the CMA). These evidences are designed to relate (support or contradict) to one or both of the models.

**Figure 1. Evidence provided to students.** Evidence #1 is from the Molecular Modeling Assessment, Evidence #3 is from the Classical Modeling Assessment.

In the MMA, students developed a model for the hypothetical skin disorder DEB while in the CMA students developed a model for FHD, a disorder that causes an irregular heartbeat. In the MMA, the first model postulated that a gene codes for a new protein, separatin, that breaks down the skin layers (Figure 2). This gain of function of an entirely new protein is not biologically plausible. The second and correct model postulated that a mutated gene lead to a non-functional protein, connectin, that can no longer hold the skin layers together. This loss of function is correct; the mutated gene will lead to a protein with a different structure and the disorder.

**Figure 2. Explanations presented in the MMA.**

In the CMA, students were presented with two competing explanations (Figure 3) about the mode of the disorder. One suggested a recessive mode in which two parents could be healthy carriers of the disorder resulting
in a 0.25 likelihood that their children will have the disorder (which is not the correct explanation based on the provided evidence). While the other, incomplete dominance, explained that individuals can be healthy, mildly or severely sick, leading to three different versions of the trait (phenotypes). The key distinction is the presence of three phenotypes, which can only be explained by the incomplete dominance explanation. This is supported by evidence which described that FHD has three phenotypes and therefore must be caused by incomplete dominance. In both cases, once students were presented with all six pieces of evidence and the models, they then responded to the prompt: "Which do you think is the better explanation of DEB, the separatin explanation or the connectin explanation? Write at least four (4) detailed reasons for your answer. Write to someone who may disagree with you, but who has not seen the evidence."

<table>
<thead>
<tr>
<th>Explanation #1: Incomplete Dominance</th>
<th>Explanation #2: Recessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Children receive one allele from each parent. The combination of alleles that they end up with determines whether they will be sick.</td>
<td>1. Children receive one allele from each parent. The combination of alleles that they end up with determines whether they will be sick.</td>
</tr>
<tr>
<td>2. Individuals with [AA] alleles are healthy and do not show any symptoms.</td>
<td>2. Individuals with [AA] alleles are healthy and do not show any symptoms.</td>
</tr>
<tr>
<td>3. Individuals with [Aa] alleles are moderately sick. They show symptoms later in life.</td>
<td>3. Individuals with [Aa] alleles are carriers of the disorder, but are not sick themselves.</td>
</tr>
<tr>
<td>4. Individuals with [aa] are severely sick and become sick at a much younger age.</td>
<td>4. Individuals with [aa] are sick and show symptoms of the disorder at a young age.</td>
</tr>
</tbody>
</table>

Figure 3. Explanations presented in the CMA.

We reported data from students who completed the argument for both the MMA and CMA assessment (n = 271). Each assessment was given to students following five weeks of inquiry-based genetics instruction. Therefore, the MMA was given following five weeks of molecular genetics instruction, and the CMA was given following five weeks of classical instruction. Students were given both sets of instruction, though half of the students (n = 144) were first given the classical form of instruction before switching to the molecular units, while the other half (n = 127) were given the molecular instruction first before switching to the classical units. All students completed both curricular units. Any students that did not complete both assessments were removed from data analysis.

Analysis
In order to assess how students constructed their arguments we coded along two dimensions; (1) structural and epistemic dimensions of evidence and models, and (2) the type and quantity of content knowledge. In terms of structural components, we looked for a claim, the amount of evidentiary backing for a provided claim, and the relationship between the evidence and claim (reasoning) (Krajcik, McNeill, & Reiser, 2008; McNeill & Krajcik, 2011). We defined using evidence as describing a connection between a specific piece of evidence to one of the models. A student response could be: scientists injected DEB patients with a working connectin protein and 80% of them got better and their skin didn’t produce blisters when rubbed anymore. Which proves that a healthy connectin was missing all along. This student described a piece of evidence (wherein scientists provided patients with a connectin-containing medication that should repair damaged skin), and connects this back to one of the models. Reasons are defined here as a justification between the model and evidence that leverages scientific knowledge (McNeill & Krajcik, 2011). Understanding how students used evidence within their arguments helps to position how students construct their argument and what was attended to during argument construction.

Students can further explain the evidence, describe the relationship between the evidence and the model, and explore how the evidence either supports or refutes a model. In terms of epistemic dimensions, we analyzed whether students critiqued the evidence and/or the models and the relationship between these features. These included such responses as: incomplete dominance is very confusing, its [SIC] not give a straight forward answer. The recessive explanation clearly explains whether it's a dominant or recessive disorder.

Content knowledge was coded as any articulation of domain knowledge that went beyond what was stated in the evidence or models. We coded prepositions and categorized how the content related to aspects of the assessment (Ruppert, Duncan, & Chinn, 2017). For example, a student during the MMA who stated the codons show a mutation which must code for a mutated gene, discussed the concept of codons. This student is describing a piece of evidence in the MMA. Three independent coders coded 36% of the CMA data and 32% of the MMA data. Interrater reliability was 91% and 87%, CMA, MMA, respectively. All disagreements were discussed.
Results
In order to understand how the assessments differed in eliciting evidence about students’ ability to write evidence-based arguments use of content-knowledge, we first analyzed dimensions of evidence use, epistemic considerations, and whether or not students used core ideas from instruction in their arguments. We discuss the features of student arguments and identify the kinds of content knowledge that students provided. We then categorized the role between content knowledge, evidence, and models in order to highlight differences across the two assessments. We identify these differences in order to discuss heuristics for design of modeling assessments.

Argument features
In looking at differences in argument structure and epistemic reasoning, we found no significant differences in selection of the canonically correct model (74%, 69%; CMA, MMA), average number of evidences used (3.5 pieces), amount of reasons (36%, 32%; CMA, MMA) or epistemic reasoning (19%, 11%; CMA, MMA). However, we found a small, but significant difference by students who used 5-6 pieces of evidence with students in the MMA more likely to use most of the provided evidence. Students who used less evidence were more likely to provide more epistemic reasoning, and students who used more evidence provided less epistemic reasoning. However, in general, our students did not generate structurally different arguments across these assessments.

Role of content knowledge
We next examined how students incorporated content knowledge. In general, students brought in 32% and 16% content knowledge into their arguments, CMA, MMA. When we investigated how students leveraged their knowledge we found content knowledge was brought in specific ways within each assessment. For each assessment, two of the types of knowledge were generally not helpful in leading students towards the canonically correct model with the third being overwhelmingly helpful.

Analysis of the MMA provided examples of explanatory prepositions. In this case, one of the evidences showed a mutated DNA sequence that may impact a protein. Some students added information stating that the DNA sequence was a sequence of codons, or added information on RNA (e.g., [this]…leads to a bad RNA and that can cause the wrong protein to be made). Students also discussed how the protein causing a problem was no longer at the cell or tissue level (e.g. I believe the connectin protein can’t connect the skin layers because it isn’t there). Like the CMA, these prepositions helped students explain a concept not explored in detail in the evidence or models. These did not help students choose the correct model, and in the case of students who brought in ideas about the protein moving from or out of the tissue layers, it overwhelmingly negatively impacted their model selection. The third type of preposition that students brought into the MMA was that the production of a new protein seemed unlikely (e.g. Usually, when someone has a disorder they are missing a gene that a normal person does. Like in albinism. I haven’t heard of someone with a disorder because of something extra). This piece of knowledge allowed students to identify the key difference between the models in terms of the disciplinary distinction at play, which is the presence of a missing non-functional protein. This is only made possible by the connectin protein story.

During the CMA, students explained carriers (e.g. in class we talked about how sometimes you can have a gene but not show it) or identified that the diseased allele must have skipped a generation (e.g. Evidence #6 said that FHD does not skip a generation. So that means even if you have the correct copy of genes, the symptoms still show up just not as severe). These two uses of outside content serve as warrants for students’ arguments. Meaning, students took a concept, explained it, and made a connection between the evidence and model. We noted that while some students explained these concepts correctly, the conclusion drawn from the interpretation either led them to select the incorrect model or did not have an impact on their model selection. These two pieces of content did not help inform the underlying mechanism critical to distinguishing between the two models. The third content preposition students used was the presence of multiple phenotypes (e.g. …the way genes work normally is that you either have this or that, but with incomplete dominance you can either have one thing, the other, or something in between, like it says in evidence 4. Which leaves the recessive invalid. In terms of flowers, if you cross-breed a red and white flower- you either get red, white, or pink). This bringing in of ideas, like the aforementioned carriers or skip generation prepositions, had students making a connection between the presented evidence and the models, but was overwhelmingly helpful in distinguishing between the two models. Students could identify that a critical idea in genetics (here the presence of multiple phenotypes) could not be explained by simple dominance (which has two phenotypes).

While students could bring in prepositions of content knowledge in one of three ways in each assessment, in each case only one was productive in allowing students to make a connection between what they know about genetics and identify the mechanistic difference between models. Beyond that, the way in which students leveraged this knowledge was different. In a simplified Toulmin argument (Toulmin, 1958), an individual makes
a claim, and must back that claim with grounds. Students in these assessments made a claim when they selected a model and provided grounds in the form of evidentiary support. In the CMA, students described how the phenomenon had three phenotypes and connected the evidence that demonstrated this (one evidence piece showed a ratio of very sick, healthy, moderately sick individuals) with the model. This is leveraging knowledge as a warrant—making a connection between evidence and model (Figure 4, right). In the MMA, students who brought in content knowledge did so as backing—providing substantiation for either the model or evidence (Figure 4, left). This difference indicated that students leveraged knowledge in specific ways in each assessment. In order to identify the affordances and constraints of each assessment we next investigated the kinds of approaches students took when structurally arranging their arguments.

Figure 4. Student uses for content knowledge in arguments.

**Approaches to argument structure**

In both assessments we noted that in general students took one of two approaches—either their arguments were structurally sound and used evidence but not content knowledge or they used content knowledge but tended to use less evidence. A smaller proportion of students combined these approaches and used both. We first detail students who used one approach.

Students were expected to make connections between the models and evidence on the basis of content as well as evaluations of how the models and the evidence related in terms of amount of supporting evidence or the quality of the evidences that provided support. These two approaches should have happened in tandem and not at the expense of one another as the purpose of our assessment was for students to demonstrate their understanding of content through the scientific practice of argumentation. However, we noted that students were more likely to make connections between the evidence and models (argue) instead of using their domain knowledge, which misses the disciplinary point of writing arguments in science. Students should feel compelled to take both of these approaches in order to have a well-founded argument. We provide examples of students who took each of these approaches in Figures 5 and 6. Student A produced a coherent argument that contains the structural and epistemic features of arguments but does not contain any outside content knowledge. This student refutes the competing model based on the quality of evidence that supports it and then shows how two pieces of evidence relate to their chosen model.

Student A:

I believe the better explanation is the connectin explanation. I believe this is true partly because the only piece of evidence to the contrary would be Evidence #4 wear a boy was given medication to break down the separatin protein by his mother. While this piece of evidence is valid this is one boy who was given something found on the internet, it could have been anything. Another reason I believe this explanation is correct is because evidence 6 is a scientific study in which scientists injected 25 patients with connectin and their skin improved. Thirdly evidence 5 is another scientific study that shows that affected people are missing.

Conversely, Student B (Figure 6) relied entirely on content knowledge and critiqued the validity of one of the models without the use of evidence. This is apt as the student supplies content knowledge but this student ignores the relationship between the models and evidence. This student does note that a gene coding for a new protein is something unusual but does not identify the distinction between the two models (that is—the presence of a missing protein). If students cannot see the disciplinary issue at hand they will not attend to it, as seen with Student A. Or, if in the case of Student B, they focus instead on the content underneath the assessment, they will...
not attend to the models. One of the key aspects of this assessment was tasking students with need to attend to both the content and the argumentation practice.

Student B: People who have DEB have a mutated gene called connectin. The gene results in a non-functional connectin protein. The non-functional connectin protein doesn’t form fibers that connect the dermis to the epidermis. The separatin explanation says that there is a new protein for DEB. Since when does a gene code for a new protein?

Student C: I believe explanation #1 is the better explanation. I believe this mostly because of evidence #4. In evidence #4 scientist took a direct survey on 115 patients with the disease. The result was out of all 115 only 12 patients between the ages of 0-3 were severely sick. While 103 were moderately sick. Now red flags start flying with this piece of evidence. The biggest one however, is in explanation #2. It talks about Dominant and Recessive genes. You can only have two phenotypes with this kind of inheritance yet in Evidence 4 there’s people who are sick, and there are people who are moderately sick. Also of course there are the people who are healthy. Three distinct phenotypes. The second biggest flag is that Explanation #2 there is no talk about someone with a moderate case of the disease. Yet in evidence #4 the majority of the people have the moderate case. Making Explanation #2 invalid.

Discussion
We presented data from arguments generated during modeling assessments. We noted similarities in terms of students who selected the canonically correct model, average number of evidences, reasons, and epistemic reasoning. Students in the MMA were more likely to use all of the evidences, however, these same students brought in half the amount of content knowledge to support their arguments as compared to students during the CMA. This difference represents a critical distinction in how students responded to the assessments. The CMA made the relevant disciplinary distinction between the competing models more salient to students. The CMA afforded students the opportunity to bring in more content knowledge by providing nonspecific mechanistic models. While these models presented a mechanism for a disorder, the steps included in the models were not based around the disorder itself and instead focused on describing the mechanism of each inheritance pattern. Students in the CMA not only brought in more content knowledge but leveraged it as warrants. Conversely, the MMA provided mechanisms that were contextualized making the underlying difference less distinct. Students who brought in content knowledge focused on the context of what the mechanism meant and provided further backing for evidence or the models but not make a connection.

Heuristics
In thinking about our findings in more general terms, we have tried to extract implications in the form of heuristics that can guide the design of assessments that can elicit evidence about student understanding in regards to both content knowledge and scientific practices. This first heuristic is to make disciplinary distinctions between models explicit. As has been stated previously, a goal in the NGSS and elsewhere is for students to demonstrate their competency in the context of some practice. This is possible if students can identify the content under investigation and can apply that in order to identify differences between models. The disciplinary goal has to be clear enough
to identify and argue about. If students cannot make a clear distinction between the models in terms of the content they will apply what they understand about the relationship between models and evidences in terms of the quality of those pieces. And while these arguments may be coherent they do not fulfill the goal of having students demonstrate their knowledge simultaneously with practice. However, the distinction cannot be too overt, as this runs the risk of students too easily identifying what the context is, and focusing on the content. In short, students should be able to see what the issue is at stake but this issue should not be made so overly accessible that students are easily provided the solution. We recognize that this is no easy challenge, and only intend to imply that the more explicit the distinction the more likely students will attend to disciplinary issue at the heart of the assessment.

A way to help make the distinction explicit without losing nuance is in the presentation of models itself. In short, the second heuristic is to decontextualize mechanistic models. The mechanisms that underlie each model must be explicit but should reinforce the disciplinary content under question by making the content more prevalent than the story used to convey that content. Looking at the two sets of models presented here, students were overwhelmed by the details to the point where it obscured the overall context in the MMA. The models were explanatory but far too contextualized for students to find what to attend to. The models presented in the CMA presented decontextualized mechanisms that allowed students to identify the disciplinary distinction. It perhaps would be better if these models were even more decontextualized so students may bring in a higher percentage of content knowledge. However, getting students too involved in the details so that they miss the overarching situation is equally problematic. Moving towards the production of mechanistic, decontextualized models that make their disciplinary content distinct presents an elegant solution for developing science assessments that could elicit student responses that integrate science content and practice. These should make the disciplinary idea under study more salient and more engaging.

References


Acknowledgements
This research is supported by a National Science Foundation Award #1053953.
In the Hive: Designing for Emergence When Teaching Complex Systems in Early Childhood

Kylie Peppler, Naomi Thompson, Joshua Danish, and Armin Moczek
kpeppler@indiana.edu, naothomp@indiana.edu, jdanish@indiana.edu, armin@indiana.edu
Indiana University

Abstract: This Design-Based Research study explores the iterative design decisions for a participatory simulation created to teach children in grades K–2 about complex biological systems. In this simulation, children assumed the role of honeybees, whose job was to collect nectar from flowers and bring it back to a hive, learning about the social nature of honeybee colonies (e.g., the need for honeybees to communicate the location of nectar via a “waggle dance”). As we designed the simulation, we iteratively analyzed the ways that children in a range of low-tech and high-tech conditions engaged in creating a nonverbal system of communication. Findings suggest that the spatial layout of the simulation directly influenced the game’s difficulty levels and subsequently impacted both the nonverbal communication and the learning outcomes of the activity. This paper presents the iterative design cycles and outlines implications for studies seeking to design participatory simulations for young children.

Introduction

Systems thinking has been identified as a useful and necessary topic to integrate into K–12 education (see Hmelo-Silver & Azevedo, 2006; Jacobson & Wilensky, 2006), operating on the belief that teaching systems thinking at an early age can help the next generation address the complex nature of an interconnected world. However, the majority of learners do not fully understand these concepts on a deep level (Resnick & Wilensky, 1993). Seeking to address these gaps as they pertain to young children’s (K–2) understandings of complex systems in science, we developed a participatory simulation (Colella, 2000), named BeeSim, that provides a first-person look into the complexity of bee colonies, in particular the interconnected challenges of collecting nectar (Peppler, et al., 2010). Similar to role-playing games, participants in a participatory simulation reenact the roles of single elements within a system, enabling them to forge personally meaningful understandings of their element’s specific behaviors, as well as its role in a greater whole (Colella, Borovoy, & Resnick, 1998). By situating imaginative exploration into a suite of elementary science curricula, BeeSim is designed to help young learners engage with complex systems-thinking concepts through play and technology (Peppler, et al., 2010).

One of the primary challenges of designing a participatory simulation for this purpose is finding age-appropriate ways to engage whole classrooms (or larger) in complex systems thinking without stifling young children’s fluid, imaginative play. Driving our design of the game was the investigation into which design features lead to the emergence of systems understanding, as well as robust and productive communication in the hive between game participants. Through a series of pilot tests and design cycles, BeeSim underwent a number of transformations, both in terms of the components of the game (e.g., puppets representing the biological systems) as well as the rules of gameplay (Thompson, Peppler, & Danish, 2017) to maximize learning and engagement. This paper analyzes the iterative design cycles of the participatory simulation, particularly the ways that design decisions impacted children’s learning of complex systems. Close attention is paid to the infusion of technological elements (e.g., electronic textiles and position tracking) into the game and how their presence changed the foundation for learning.

Complex systems thinking, biological systems, and children

A system is recognized as “complex” when the relationships within it are not obvious, and the individual elements of the system give rise to new overall properties that are difficult to see or explain (Hmelo-Silver & Azevedo, 2006). This is especially true in biological systems, where individual organisms may act in ways that seem counterintuitive when compared to the behavior of the system as a whole. For example, individual honeybees spend a considerable amount of time “dancing” to communicate nectar location to other bees in the hive. However, this behavior gives rise to faster and more efficient nectar collection for the hive as a whole. Young children, however, tend to assume this time spent dancing is wasteful (Danish, 2014). This surprising interaction between levels (Wilensky & Resnick, 1999) in the system is known as emergence.

Much of the work around systems thinking education has been through biological systems; much thought has been given to teaching biology, or life science, to young learners, as it is a topic children are
familiar with and curious about. For example, Hmelo-Silver has often studied children’s understanding of aquatic and respiratory systems (e.g., Hmelo-Silver, Marathe, & Liu, 2007), while Wilensky has looked into large ecologies involving wolf, sheep, and grass (e.g., Wilensky & Reisman, 2006). Although these studies were not conducted with children in our target age range, their findings help us see the benefits of exploring complex systems through biological systems. Wilensky and Reisman (2006) found that simulations employing agent-based models helped students think more deeply about complex systems and relate the agent-based occurrences to the aggregate level occurrences. Games are especially powerful because they allow children to take on new perspectives through play, supporting productive learning (Peppler, Danish, & Phelps, 2013).

This project emerged from a longer history of research using bees as a vehicle for teaching complex systems (Danish, 2009; Danish, Peppler, Phelps, & Washington, 2011). Most notable of these is BeeSign, a 2D computer-based program that offers a third-person perspective on biological systems (Danish, 2009). The overarching goal of adapting BeeSign to a participatory simulation was to offer a play-based, first-person perspective on the behavior of bees with a focus on the challenge of finding nectar, engaging in the design of nonverbal communication (i.e., how bees convey the location of flower nectar to other bees), and other variables, such as nectar quality, nectar depletion, and a limited flight range. We aimed to develop a high quality first-person participatory simulation for two reasons: (1) to engage entire classes of children in a single systems thinking activity, which would better reflect the collective behaviors of biological systems; and (2) we sought to create two parallel first- and third-person conditions (or at least to the extent possible) to disentangle the differential contributions of first- and third-person experiences on learning of complex systems in early childhood. The key challenge of the latter goal was to find ways to control for the overall difficulty level, depth of systems thinking concepts engaged in the activity, and overall entertainment value of the game play for each of these conditions.

Methods

Using a Design-Based Research paradigm (Design-Based Research Collective, 2003), the four most recent iterations of BeeSim discussed here were tested in two mixed-grade (grades 1-2) classrooms and one first-grade classrooms at different suburban, midwestern elementary schools; design cycles 1 and 2 at one school, cycles 3 and 4 in two comparable schools in the district. Three teams with seven to nine students in each engaged in cycles 1 and 2 (fifteen boys and eight girls), while two teams with between ten and twelve students in each (twelve boys and ten girls) participated in cycles 3 and 4. Each team included at least one member of each gender. It is important to note that earlier versions of BeeSim did take place between BeeSign and the four cycles discussed here, but designs were less pertinent to the current, optimized design reached in cycle 4.

Before each game began, participants were divided into two teams and were told about the objectives of gameplay: bees had to collect nectar and bring it back to the hive within a fixed time period, bees needed to communicate to each other about which flowers had the most nectar, and no bee could talk during the game, only signal the location of the nectar to other bees via a “waggle dance.” The BeeSim curriculum covers a number of aspects of the hive system, including the constraints bees face as they struggle to feed the hive, how bees communicate, and why this communication is important. In the honeybee system, unlike other systems in the animal kingdom, communication is done through a waggle dance in which returning foragers perform a series of figure-eight movements to communicate both the direction and abundance of a food source to her sister bees. The waggle dance contains quite a bit of information about the direction, distance, and type of nectar. As children learn about the waggle dance, they invent a novel waggle dance to communicate with their hivemates in game play.

Observations from each of the four major design cycles highlighted here were accumulated through a combination of videotaped game sessions, written observations and ex post facto discussions between the research team and the children participants’ homeroom teachers (who integrated BeeSim into additional activities in the classroom, including discussions and engagement with the BeeSign curriculum). Through discussions with teachers, as well as observations of how children engaged in the simulation during game play, the research team was able to determine which learning objectives and game dynamics could be optimized through refinement of the game rules and components between cycles.

Findings

Data collected from each design cycle helped to clarify the design elements of the simulation that were linked with children’s engagement with authentic biological and systems-thinking concepts. The sequence of observations, and the design iterations they necessitated, are presented in chronological order below in Table 1. Note that BeeSign has been included in this table as a reference to the starting point of these activities. As later iterations were built and refined, we often looked back to BeeSign as a point of comparison. We worked to
mirror the productive design choices where appropriate, and departed in careful and intentional ways, such as where the shift from third-person perspective to first-person perspective changed gameplay and interactions.

Table 1: Summary of the various design cycles, including key design features and outcomes

<table>
<thead>
<tr>
<th>Design Features</th>
<th>BeeSign (2D computer simulation)</th>
<th>BeeSim: Low-Tech Conditions</th>
<th>BeeSim: High-Tech Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Simultaneous Actors</td>
<td>10</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>Number of Flowers</td>
<td>2+</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Distribution of the Flowers in the Field</td>
<td>Selectively distributed around the hive to promote effective inquiry.</td>
<td>Randomly distributed in a large, outdoor space</td>
<td>Organized in a small grid of 2 rows of 3 flowers in a 2’x3’ area</td>
</tr>
<tr>
<td>Component Design</td>
<td>Computer simulation, interactive whiteboard</td>
<td>Paper flowers, eyedroppers</td>
<td>Life-size flowers, Arduino-enhanced puppets, RFID tags communicate between puppet, flowers and hive</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Quality of Dance Communication</td>
<td>n/a</td>
<td>Infrequent, Simple</td>
</tr>
<tr>
<td></td>
<td>Systems Thinking Concept of Emergence present</td>
<td>Yes</td>
<td>X</td>
</tr>
<tr>
<td>Design Principle / Takeaway</td>
<td>n/a</td>
<td>Include constraints that make key phenomena salient.</td>
<td>Include technology in ways continue to provide authentic constraints.</td>
</tr>
</tbody>
</table>

**Cycle 1: “Low-tech” condition with distributed field**

In the first iteration of BeeSim, participants carried an eyedropper for use as a proboscis (the straw-like mouth of a bee) to collect nectar from twelve laminated paper flowers. The flowers were scattered randomly around a large outdoor garden space, anywhere from 20–40 feet from the hive (see Figure 1), behind each were situated Dixie cups of colored water representing nectar. Once the simulation started, each team could set out a “forager” to search for nectar in the yard, and then were asked to nonverbally communicate the location of the most nectar-filled flowers to the other bees back in the hive using an invented “waggle dance.”

Findings from this design revealed a number of challenges for the young age group. For one, (1) the children were easily distracted by the size of the field, likely exacerbated by (2) the random placement and copious numbers of flowers, which made it difficult for children to formulate and communicate meaningful waggle dances to other members of their hive. The free structure of the game also led to (3) participants not returning to the hive in a timely manner, as well as (4) cheating by surreptitiously peering at the nectar levels of all flowers, thus undermining the importance of the “waggle dance” and weakening the emergence of biological or systems understanding. Reducing the number of flowers and size of the playing space was hypothesized as a way to lower the barriers to success, as well as insert ways for adults to moderate how participants collected nectar and communicated with the hive. From these discoveries, we might suggest a concrete takeaway for other educators designing activities for young elementary learners: include constraints that make key phenomena and
learning goals salient for learners, and identify ways to enforce those constraints. In our experience, a focus on “winning” a game can lead learners to circumvent rules and constraints if possible.

**Cycle 2: “Low-tech” condition with optimized field**

We next tested the game with the same children as before, but the total flower count was reduced to six and arranged more closely together: in a 2x3 grid approximately 20’ from the hive. Additionally, childrens’ adherence to the rules of the simulation were enforced by adult moderation. In this instance, a pair of two children from one hive were permitted in the field at any given time. To collect nectar, children indicated to an adult stationed by the flowers which flower they wanted to check, and after checking this flower their imaginary energy ran out and they had to return to the hive. Back at the hive, these students used their dance to nonverbally communicate which flower the next bee pair should check.

![Figure 1](image)

Figure 1. Field layout in low-tech cycle 2 (top) and high-tech cycle 3 (bottom) conditions.

The optimized game space and presence of adults to scaffold the rules of play resulted in consistently on-task performances from the children, including the invention of elaborate waggle dances to communicate the nectar location to team members (see Table 2). After the simulation, we observed that children’s resulting conversations about nectar collection were productive and insightful about how bees communicate within the hive. In particular, many of the students began to recognize the benefit of each bee dancing for the performance of the hive as a whole—a rudimentary recognition of the emergent properties of the system.

While the emerging analyses suggest that learning may have improved through these design changes, we identified a need to make the gameplay experience more active for all participants, especially since youth not actively “foraging” had little to do back at the hive. Furthermore, while the nectar collection was more systematic in this instantiation, having to communicate first to an adult before collecting is distal to the authentic biological model. Our next design cycle was driven by the idea that a technology-enhanced environment could communicate the biological rules of the bees more consistently, reducing the need for adults to police youth children’s game choices, enabling more youth to play simultaneously, as well as opening the door for more biological science to be integrated into bee behaviors due to a reduced emphasis on enforcing the rules. Design takeaways from this iteration center around the decision of whether and how to introduce technology to a learning experience. Here, the inclusion of technology was positioned as a way to provide constraints on students’ behavior that was authentic to the learning context.
### Table 2: Summary of “waggle dance” communication created by teams in BeeSim v2.0

<table>
<thead>
<tr>
<th>Condition</th>
<th>Hive Name</th>
<th>Communicate Direction</th>
<th>Communicate Distance</th>
<th>Communicate Quality/Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeeSim v2.0</td>
<td>Team 1</td>
<td>Hand above head if on the top shelf, at neck height if on the middle shelf, mid-chest height for lowest shelf</td>
<td>Hold up one, two, or three fingers to indicate how far along the shelf the flower is</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Team 2</td>
<td>Jump and point up for top shelf, jump and point down for bottom shelf, just jump for middle shelf</td>
<td>Hands to the left if it is the first flower on the shelf, middle if it’s the second flower or to the right if it’s the third</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Team 3</td>
<td>Thumbs up above head for top shelf, thumbs up straight in front for middle shelf, thumbs up down low for bottom shelf</td>
<td>Thumb bent to to the right for closest flower, straight up for middle flower, tilted out for the far flower.</td>
<td>Using the other hand, thumbs up if there is nectar or thumbs down if there is not</td>
</tr>
</tbody>
</table>

#### Cycle 3: “High-tech” condition with distributed field

Seeking to increase children’s autonomy and support a larger number of actors, as well as increasing biological realism in the next iteration of the game, we designed computationally-enhanced puppets and a field of uniquely designed electronic flowers into the participatory simulation. Making use of the LilyPad Arduino platform—a microcontroller board that can be stitched safely into textiles—each player used bee puppets that included an XBee 2.5 2mW wireless module multiple sets of 3 LEDs, and an RFID reader. The XBee Wireless Module allowed for wireless communication between the glove and another XBee attached to a computer via USB. During gameplay, students wearing the bee puppets could monitor through a set of three LEDs the amount of nectar currently stored on the glove, while an additional set of LEDs in the bee’s antennae displayed the amount of nectar in each flower checked. To represent the finite energy levels of bees as they travel between the hive and a flower, a tri-colored LED was used as an energy meter, moving from green to red to indicate to students when they needed to return to the hive.

The flowers incorporated RFID tags that would be scanned by the bee puppet to “check for nectar.” In addition to being more proportional in size to the bee puppets, the flowers were also diversified in terms of variety of flower types and collection methods, and were once again returned to a more random, distributed arrangement in the playspace (see Figure 1). When the RFID scanner in the bee puppet came near the RFID tag in the flower, the computer noted the time and flower ID of the collection. If the child returned to the hive before energy ran out, the total amount of nectar for the team increased by the amount of nectar currently stored on the bee. The initial rationale for returning to a greater number of flowers was that we hypothesized that having more actors in the field would necessitate more overall coordination of activity as a result of needing to communicate which source of nectar the bees should visit. Given that this was indoors and arranged around the existing furniture for a temporary installation for the class, this also led a random distribution of the flowers, which we had hypothesized as not being problematic given the new time constraints placed on the system.

The introduction of the technology led to some new affordances in the game. Not only could more children participate at a time, but the puppet itself communicated nectar storage and energy levels to the children, opening up the gameplay for the emergence of new biological understanding. For instance, in this version, children had a limited amount of time to collect and deposit nectar and a finite storage capacity. During the allotted time, a child would run from flower to flower and try to collect nectar. A child could collect one unit of nectar from any given flower and would also be informed as to how much nectar remains inside the flower.
Once the child’s nectar stomach (represented via an LED array) was filled, he or she returned to the hive and deposited the stored nectar.

Furthermore, the electronics-based version of the BeeSim game gave the instructor more freedom and better access to data than in previous incarnations. For instance, because the energy of a bee was monitored by an adult’s stopwatch in BeeSim 2.0, children often failed to consider the range to be a real constraint for bees, because it was not a real constraint for them. In contrast, the computational textile bees embedded the bees’ energy into the game in a natural and familiar manner such that the children in the role of the bees had to attend very carefully to it, or suffer the consequences (lost nectar). This resulted in far more attention to details important to understanding the system. In addition to the bee range, the computational textiles also helped to model limited amounts of nectar collection, flower variables such as random nectar depletion and the difficulty of determining if a flower has nectar without visiting it, and supported easier tracking of how much nectar was collected.

However, there were key challenges in this version of the game as well. While constraints on time, etc. were all designed to help the children reflect upon the constraints that real bees face as they collect nectar, as well as the benefits of the solutions that honeybees have evolved to these constraints (e.g., the waggle dance to convey nectar sources), it also had inadvertent negative impact on the observed quality and frequency of the communication in the hives (see Table 3). This stemmed from the infrequency of finding nectar in the field, which was due to the random placement of the flowers in the field, overall number of flowers, as well as the limited amount of time to search for nectar. For example, one bee would only have a limited amount of time in the field to find nectar, which would oftentimes be unsuccessful until the game was too close to completion, leaving very little time to use the waggle dance and for others to “listen” to and act on this information before the end of the class period (see Table 3). The subsequent video analyses of this version led us to see a pattern of play that is problematic with the distributed field of flowers (as similarly seen in cycle 1); overall complexity of the task is increased when there are more than 6 or so objects in the field to check. Debrief and reflective conversations, did work to address this complexity, but it felt necessary to simplify the nectar collection task as well. Thus, in subsequent versions of the game, we sought to reduce the overall number of flowers to help provide an easier entry into the game play and help to focus the actors on the behaviors (i.e., the waggle dance) that lead to better understanding of emergence within the system. Here, we might see design takeaways that highlight tensions between physical simulations and computer simulations. The number and type of elements in each might be different and can only be determined through testing.

Table 3: Summary of “waggle dance” communication created by teams in BeeSim v3.0

<table>
<thead>
<tr>
<th>Condition</th>
<th>Hive Name</th>
<th>Communicate Direction</th>
<th>Communicate Distance</th>
<th>Communicate Quality/Amount</th>
<th>Communicate Flower Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeeSim 3.0</td>
<td>Team 1</td>
<td>Shake body in the direction of the flower.</td>
<td>Indicate number of steps using fingers.</td>
<td>N/A</td>
<td>Purple: pinch thumb and forefinger Pink: touch face Orange: fist touching chin Yellow: mimic peeling a banana</td>
</tr>
<tr>
<td></td>
<td>Team 2</td>
<td>Numbered flowers one through six from right to left in terms of their placement in the field. Step out the shape of the number.</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Cycle 4: “High-tech” condition with simplified layout
In the latest testing of BeeSim, participants used the electronically enhanced components but the gameplay returned to the optimized number and arrangement of flowers from v2.0. Overall findings indicated that
performance on pre-to-post measures improved in every case, indicating strong learning gains from participation in this cycle. While the designs and rules of play were similar to cycle 3, new game designs emerged when the overall number of flowers was reduced, which was kept at 4–6 throughout the testing. Generally, one of the flowers was a distractor that would not produce nectar appealing to honeybees in nature. Additionally, another flower was often set to “empty,” and one more was partially filled with nectar. Although the smaller number of flowers meant that participants might be able to check multiple flowers in a relatively short amount of time, with so many of the flowers empty or nearly empty, participants needed to utilize the waggle dance to be most efficient and avoid losing time/energy at empty flowers. The flowers were placed at random locations in the classroom, both to prompt the need for unique waggle dances, and to accommodate immovable classroom furniture.

As a result of this optimized flower arrangement, waggle dances became more efficient during this round of design. Rather than attempting to convey flower color, as had occurred in v3.0, participants focused in on the quality and amount of nectar of the flowers (see Table 4). The faster or slower shaking used by Team 1 mirrored the faster or slower wagging performed by real honeybees. To convey quality, participants needed to attend to an additional piece of information, found in the antennae of the bee puppets. Lights here would flash very quickly for high quality flowers, and more slowly for low quality flowers. Participants in v3.0 were less likely to notice this information as there were many flowers to choose from, and quality became less consequential. This iteration of game play revealed more frequent and more complex communication within the hive sooner in the game play, which allowed for participants to experience and note emergent behaviors in the hive (i.e., that you can more efficiently collect nectar when you communicate where to go to find nectar with your hivemates). This leads to a final design takeaway: it is crucial to check in through the iteration process to ensure core learning goals are not overshadowed by design changes.

Table 4: Summary of “waggle dance” communication created by teams in BeeSim v4.0

<table>
<thead>
<tr>
<th>Condition</th>
<th>Hive Name</th>
<th>Communicate Direction</th>
<th>Communicate Distance</th>
<th>Communicate Quality/ Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeeSim 4.0</td>
<td>Team 1</td>
<td>Shake body in the direction of the flower</td>
<td>Hold hands close together, medium distance, or far apart</td>
<td>Faster shaking to indicate higher quality, slower shaking to indicate lower quality</td>
</tr>
<tr>
<td></td>
<td>Team 2</td>
<td>Shake body in the direction of the flower</td>
<td>Indicate number of steps using fingers</td>
<td>Rub stomach and nod head as if &quot;yummy&quot; to indicate higher quality, give thumb's down to indicate lower quality</td>
</tr>
</tbody>
</table>

**Discussion**

Through iterative cycles of development, the various versions of BioSim represent a shift in a number of the design features, enabling and constraining the children's activities via the designs and computational affordances of the wearable computers to make visible more aspects of the system. In short, the overall designs of the system were optimized for this age group when the (a) time to discover nectar was constrained by the puppet design and (b) the overall number of flowers in the field was reduced to minimize the number of possibilities for this age group Such findings provide insights into the complexity involved in promoting systems understanding for this age group, and they help us understand the role that spatial design plays in capitalizing on the affordances of learning materials, but they could have gone unnoticed had the research focused only on the summative impact of the intervention.

In sum, these cycles of development revealed how technology could facilitate the implementation of robust game rules, highlighting the importance of communication within the hives. Future participatory simulations designed for young children can build on the design takeaways outlined here. We see them as emerging principles that could guide the design of other types of contexts beyond biological systems and social insects. Furthermore, achieving one of the primary goals of the simulation (i.e., first-person perspectives on the
bee’s role in the greater system), the technological enhancements made to the BeeSim simulation in the later iteration of the game facilitates changes to the game rules that simulate an experience closer to that of real honeybees, which could be skinned and used in other applications to model parallel systems, such as blood circulation and/or army ants foraging for food. The findings in this paper serve to highlight the importance of individual design elements within a system and how these elements work together to shape the system as a whole.

References


Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1324047 awarded to Kylie Peppler, Joshua Danish, and Armin Moczek. 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 National Science Foundation. Thank you to Janis Watson, and the many teachers and students who made this work possible.
Becoming Facilitators of Creative Computing in Out-of-School Settings

Ricarose Roque, University of Colorado Boulder, ricarose@colorado.edu
Rupal Jain, Woodrow Wilson Academy of Teaching and Learning, jain@woodrow.org

Abstract: Educators in informal learning environments are often called “facilitators” to highlight the role that they play in guiding youth interests rather than prescribing or directing their activities. This paper presents research on the development of adult facilitators in out-of-school programs that engage youth in creative computing experiences, where youth create, design, and build with technology. Facilitators are typically volunteers recruited from the local community, schools, and companies with varied backgrounds with youth, education, and technology. We examined the experiences of adult facilitators in an intergenerational learning program that supported youth and their parents to learn and create with technology. We describe three case studies of adults who developed multiple practices and roles and discuss what supported their development as facilitators. This study has implications for how informal learning organizations and programs can support facilitators from varied backgrounds to support diverse learners in creative computing.

Introduction
In the last decade, researchers, policymakers, and industry leaders have recognized the importance of supporting youth in learning to code, or creating and designing with computing, to become full and empowered participants in our increasingly digital society. Many efforts have emerged in out-of-school settings, such as after school clubs like CodeClub, international challenges like Technovation, and libraries like Ready to Code.

To support these experiences in informal learning spaces and to engage more diverse groups in computing, adult facilitators can play important roles in developing welcoming spaces that enable youth to learn with computing (Vossoughi, Escude, Kong, & Hooper, 2013). In this paper, we focus on the experiences of adult facilitators in out-of-school settings designed to engage youth in creative computing, or making, designing, and tinkering with technology. Adult facilitators may be students, professionals, or community volunteers. They might help for a few days or remain for a few years. Some might have extensive experience with computing as engineers or technology designers, while others might have limited experience. Some might have backgrounds as educators with experience working with youth, while others might have very little. Facilitators might range in age from teenagers to retirees. These adult facilitators are often recruited and trained by staff from informal learning organizations. With these varying backgrounds and circumstances, staff in informal learning spaces face challenges in supporting facilitators to develop the practices that can meaningfully engage diverse learners.

While many studies have examined the experiences of youth in these programs, relatively fewer studies have examined adult facilitators and what supports they need to facilitate empowering and enriching experiences for learners of many backgrounds, different interests, and vastly different expertise with technology. Past studies have primarily focused on the practices of facilitators, such as surfacing learners’ interests, providing encouragement, guiding rather than directing, and deepening their engagement (Gutwill, Hido, and Sindorf, 2015). Other studies of adult caretakers such as parents and mentors highlight the different roles they can play such as teacher, collaborator, and learner (Barron, et al 2009; Nacu et al, 2016; Kafai et al., 2008). The identification of these roles and practices are important to help define what facilitators can do, but how do students, professionals, and volunteers learn to take on these roles?

In this paper, we take an initial step to more closely examine the development of adult facilitators and what supports their development of the practices and roles highlighted by these past studies. We present the experiences of facilitators participating in an intergenerational learning program called Family Creative Learning (Roque, 2016). Within this program, facilitators support kids and parents to create and learn together using creative technologies like the Scratch programming language. Facilitators were recruited from local universities, companies, and from the staff at a community organization hosting the workshops. Facilitators had mixed backgrounds as former classroom teachers, software developers, and community organizers.

Our analysis of facilitators’ experiences builds on sociocultural frameworks of learning, where learning is embedded in shared activities and involves taking on a variety of practices and roles that change over time (Lave and Wenger, 1991; Rogoff, 1994). We ask (1) what kinds of roles did facilitators play as they engaged families in creating and learning with technology? (2) what supported facilitators in the development of these roles? We focus our study on facilitators in this family program, using collected ethnographic data to describe
facilitators’ development. We argue that for facilitators to develop roles to effectively support learners, they must also engage in the process of learning themselves. The adults in our case studies were continually learning as they encountered challenges, responded to learners’ interests and needs, and tried to support families.

**Background**

Educators in informal learning environments are often called “facilitators” to highlight the role that they play in guiding youth and supporting youth interests rather than prescribing or directing their activities (National Research Council, 2009). Supporting such learning experiences require facilitators to negotiate a variety of practices such as knowing when to step back and observe or when to intervene to provide encouragement and feedback. Studies of facilitators in spaces such as makerspaces and afterschool programs reveal the depth of their practices to spark interest, sustain engagement, and deepen learning trajectories (Gutwill et al., 2015).

Facilitators must carefully balance offering enough support so that learners can actualize their ideas with maintaining enough distance so that learners truly drive the experience for themselves. Striking the right tenor can be challenging, forcing facilitators to question their own interactions constantly: Should I intervene yet or allow that girl to be frustrated for a few more seconds? (Gutwill et al., 2015, pg. 161)

Additionally, facilitators play critical roles in engaging non-dominant youth and supporting them to access new and emerging opportunities with technology (Barron, Gomez, Martin & Pinkard, 2014). Facilitators help to create an environment of intellectual and emotional safety and build on the “funds of knowledge” and “repertoires of practice” of learners (Vossoughi et al., 2013). Facilitators can play important roles positioning learners as creators, pushing back against narratives of being consumers of technology (Barron et al., 2014).

While facilitators enact important practices to support creative and inclusive learning experiences, facilitators themselves may not necessarily be professional educators with a background in supporting diverse youth to engage with digital opportunities. Staff at informal learning organizations may recruit local STEM professionals, students, and community volunteers with a variety of backgrounds and experiences with technology and education. Some facilitators may engage with an informal space for a few weeks to years.

Such varied backgrounds and participation patterns pose challenges for training and ongoing professional development. In a study of three makerspaces, Litts et al. (2015) described the challenges of makerspaces in equipping facilitators with the necessary skills and training. The disciplinary backgrounds and past experiences of facilitators tended to influence the moves and decisions they made as facilitators. For example, facilitators with a background building bikes might facilitate bike building activities. Those with backgrounds supporting physical making might lean towards supporting activities and trajectories in the physical realm — limiting learning opportunities with digital media.

Researchers of informal spaces emphasize the need to better understand how facilitators develop and what supports their development (McNamara, Akiva, Wardrip, & Brahms, 2016). This paper focuses on these questions in the context of a family learning program.

**Facilitators in Family Creative Learning**

In this paper, we explore the role of facilitators and what supports them to take on different roles within a community-based program called Family Creative Learning (FCL). FCL invites families to learn together using creative technologies (Roque, 2016). FCL has five workshops and are held in a community center once a week for two hours each. Each workshop is divided into four parts: Eat, Meet, Make, and Share. In Eat, families have a meal from a restaurant, allowing all families and facilitators to eat with each other. In Meet, facilitators check-in separately with parents and children to talk about their experiences in the workshops. In Make, parents and children engage in design activities with the Scratch programming environment and Makey Makey invention kit. In Share, families talk about their projects to other families and receive feedback and questions.

Unlike instructionist learning environments where there is a central instructor and pre-determined activities, the design of FCL draws on constructionist traditions of learning, which argue that people learn best when they are building things that are personally and socially meaningful (Papert, 1980; Kafai, 2006). Constructionism builds upon constructivist traditions that knowledge is not something that is transmitted or acquired, but something that is actively constructed through experience (Piaget, 1976). As people build projects, they build ideas. To be personally meaningful, the design of FCL invites families to build on their diverse “repertoires of practices” and “funds of knowledge” (Gutiérrez & Rogoff, 2003; Moll, Amanti, Nef & Gonzalez, 1992). To be socially meaningful, the design of FCL has also leveraged learning theories that emphasize the
social aspects of learning (Brown, Duguid, & Collins, 1989; Lave & Wenger, 1991). Families are encouraged to work together as well as interact with other families participating in FCL.

Prior to the workshops, facilitators met as a team to become familiar with the tools, activities, and facilitation. Both authors met with facilitators and highlighted the role they play in supporting families to engage in personally and socially meaningful experiences with computing. They shared the Family Creative Learning Facilitator Guide (Roque & Leggett, 2014), particularly the Facilitating Fundamentals, which included practices such as “build relationships and trust”, “ask questions rather than giving answers,” and “surface their interests.” For example, to build relationships and trust, facilitators are encouraged to eat with families and focus on a few families rather than the whole room so that facilitators can form deeper connections with families. Additionally, the authors prepared some facilitation dilemmas for facilitators to discuss such as what to do if someone asks a challenging question or what to do if someone becomes frustrated with the making process.

During the workshops, facilitators supported the workshop implementation and worked with families to help them with their projects. Immediately after a workshop session, facilitators met to debrief for 30 minutes to discuss what went well, what questions they have, or things that could be improved or challenging interactions they witnessed. Between the workshops, the facilitators met again to consider their reflections from the previous workshops and to discuss changes or strategies to implement in the next workshop.

Studying the experiences of facilitators

Participants
In this paper, we focus on facilitators who participated in the design and implementation of FCL workshops conducted in the Spring of 2015 at a public housing community center in an urban community in the northeastern United States. Six facilitators facilitated and six families participated in the workshops (13 family members). Facilitators were recruited through local universities and through volunteers and staff from the participating community center. Two facilitators were graduate students pursuing a one-year, professional masters in education, one facilitator was an engineering undergraduate student, one facilitator was a software developer, and two facilitators were staff from the community center. Students were recruited by emailing students lists at two local universities. Facilitators’ backgrounds with computer programming ranged from limited exposure to Scratch programming in a class to using programming as part of their educational and professional background. Facilitators were not paid to participate in the FCL workshops. Roque served as a facilitator during this program, while Jain served as a facilitator in past FCL programs.

Families were recruited from the community, with kids between 7 to 13 years old and parents between 35 to 82 years old. (We use parents loosely to mean any adult caretaker such as grandparents and family friends.) Five of the six parents were women. All parents were immigrants from countries in Latin America. All kids qualified for free/reduced lunch in school. All families created and shared a project at the final showcase.

Data collection and analysis
While the data in this study focuses on facilitators, this data is also part of a larger qualitative study using ethnographic and case study methods to examine the learning experiences of families within the context of computing (Roque, Lin, & Liuzzi, 2016). We used individual interview data to understand child and parent interactions and perceptions of facilitators. To understand the experiences of facilitators, we used multiple methods of data collection. During the workshops, we collected photo, audio, and video documentation of facilitator and family interactions. We audio-recorded, and later transcribed, team reflections after the workshops. Between workshops, facilitators wrote field notes to describe the experience from their perspective. We also documented team meetings, where we discussed plans for the next workshop based on reflections from the prior workshop. Three facilitators (the two staff members and the software developer) were unable to participate in writing field notes and meetings between workshops due to their work schedules. After the program ended, we conducted 60-90 minute semi-structured interviews with the six facilitators. Prior to the interviews, we asked facilitators to reflect on three to five moments that mattered to their experience. In addition to asking questions about these moments, such as why they chose them and what they took away from those moments, we asked them questions to surface their motivations, their facilitation challenges, and the strategies they developed. These multiple methods of data collection before, during, and after the program allowed us to triangulate their experiences as well as capture their development as facilitators over time.

We used grounded theory strategies (Charmaz, 2006) to uncover experiences of development or change by examining what facilitators did, how they interacted with families, and what they said about their experience. In particular, we examined the moments that facilitators identified in their interviews — moments where they learned, where they encountered a dilemma, or where they shifted their thinking. We then used field notes,
facilitation workshop debrief sessions, and audio and video documentation to further examine these particular moments of learning or change. We used roles identified by other studies of adult caretakers, such as teacher, collaborator, learner, and audience roles, and we used the grounded approach to uncover additional roles and practices (Barron et al., 2009; Nacu et al., 2014; Brahms & Crowley, 2016; Kafai et al., 2008). Both authors participated in data collection, transcription, coding, and analysis. We met weekly during a seven-month period to discuss data analysis, making sure that each analysis was approved by both team members.

Of the six facilitators, we focused on three facilitators to develop case studies to highlight the ways that facilitators took on varied roles and what supported them to take on these roles as they supported families. These three facilitators were selected to showcase different experiences of development, but they are no means representative of all facilitators. While we focus on these three, the other facilitators in the workshop are embedded in their experiences since the facilitators interacted in several ways throughout the workshop.

**Results**

In the following sections, we present three case studies of facilitators’ experiences that reveal how they were able to shift roles, practices, and perspectives to support and guide families. We focus on their practices (e.g. sharing reflections, asking questions), their interactions with families (e.g. making suggestions, co-creating), and their perspectives on themselves and the families (e.g. feeling insecure about their expertise, seeing how open parents can be). We use pseudonyms to refer to the facilitators and families.

**Maria: Developing a tinkering mindset**

Maria, a former classroom teacher and graduate student at a local university, was looking forward to helping with the workshops. As a native Spanish speaker, she was especially interested in helping Spanish-speaking families. However, she worried how helpful she could be with her limited experience with computing.

During Workshop 1, Maria helped parents get started with Scratch, who were engaging in an activity to animate the letters of their name in Scratch. She helped one parent, Andres, who recently immigrated and primarily spoke Spanish. He was having trouble getting started because the activity handouts were in English. She showed him how to add letters, and after watching him add two letters, she continued walking around the room. When she came back, he proudly shared his project with Maria. She asked him what he wanted to do next. He mentioned that he wanted to add sounds because his nephew liked music. Maria pointed out the sound features in Scratch. During the debrief with other facilitators after the workshop, Maria shared how surprised she was to see how open the parents were to trying things out and being vulnerable to not knowing. Meanwhile, she shared her desire to continue improving her experience with Scratch.

During Workshop 2, Andres worked with another parent, Julia, and they called over Maria for help. They were trying to program their project to play a drum sound whenever the space key was pressed. Maria saw that their Scratch program looked correct and became puzzled as to why their project was not working. She called over another facilitator, Alex, who was a software developer and had past experience running Scratch camps. He also recognized their blocks, but wondered if the Scratch editor did not have the keyboard focus, which allows the Scratch editor to notice keypresses. He asked the two parents to click on the Scratch editor and after they pressed the spacebar again, the project worked. Soon after this interaction, they called on Maria again to help them make their sprites dance. Maria was unsure and she called on another facilitator to help. Maria felt frustrated that she could not immediately help the two parents.

Maria shared these feelings of frustration during the team debrief. Another facilitator, Pia, who had past experience facilitating Scratch workshops, responded, “I don’t think you should feel that it’s all on you [to know]... You don’t have to know. What you have to focus on is modeling the tinkering and exploration and the experimenting. I think it’s the attitude that’s more important.” In a post-workshop interview, Maria noted how Pia’s comment changed her mindset and approach as a facilitator. She became less self-conscious and instead focused on showing participants how to go about realizing their ideas. When she was unsure, she would ask herself: “What could I do? Where could I go? Who could I ask?” rather than worrying about her lack of expertise. She added, “That's big for me because before that I was a teacher and a teacher always has to know.”

Maria’s experiences also shifted her notions of what it meant to be a teacher. Instead of being someone who “always has to know,” she shifted her role to model how to tinker, experiment, and learn from challenges. Additionally, in her post-workshop interview, Maria noted that her perspectives on parents had shifted. When she was a teacher, her interactions with parents were often about homework and discipline. “For the first time, I connected with a parent. That never happened before for and that was powerful. They're so curious and full of joy. I never saw that side of parents before.”
Case analysis
Maria’s experience is not unlike many of the other new facilitators who had limited experience with computing. Through pre-workshop facilitation meetings, Maria gained some basic experience with Scratch. At the start of FCL, Maria took on a teacher role (which she was familiar with as a former classroom teacher) and she shared her expertise with Andres. However, as families continued to advance and their ideas became more complex, Maria’s knowledge of Scratch reached its limits and she felt frustrated at her inability to fulfill her teacher role.

The facilitation team was an important resource for her to help her answer questions from families and to shift her role in the workshops. Maria pulled on different members of the team who had varying expertise with Scratch, such as Alex, to aid in her facilitation. Rather than being someone who had the expertise, Maria realized that what was important was her ability to model how to learn through tinkering and exploration as well as through pulling on different people in the room.

Sam: Helping them pursue their ideas
Sam, a former special education teacher and graduate student at a local university, had past experience helping his students learn with technologies, such as robots, construction kits, and virtual worlds. However, he had little experience working with families and looked forward to learning how he could engage parents and kids.

He met Julia and Eric, a mother and son, early in the workshop series, and noticed that both were shy and quiet. He sometimes had to lean in to hear them speak. However, once he sat down and engaged with them one-on-one, he learned that Eric liked building things, and had just started to play with a new computer that Julia got from participating in computer classes at their community center. When they began their family project, Sam watched Eric and Julia as they quietly discussed their project and Eric drew their project idea. Sam checked in with them and learned about their rollercoaster idea. However, it was near the end of Workshop 3 and Sam was concerned about what Julia’s body language communicated:

She was directing Eric more, and she would be like, "Do that." She was looking a lot at what [another family] Estella and Carlos were doing. And I don't think she was competing with them, but I think that she wanted to produce something that she was proud of. And they [Julia and Eric] definitely liked their idea. They wanted to make the roller coaster.

He watched as they tried to find a rollercoaster image through Google. After searching for some time, he stepped in and suggested drawing their rollercoaster on Scratch showing them the paint editor in Scratch. However, as they got started, he noticed they had trouble working the mouse, frequently right-clicking instead left-clicking. They were unable to make much progress and by the Share portion of the workshop, and he felt a low energy from them as they described what they tried to do.

In Workshop 4, Sam decided to work more closely with Julia and Eric, asking them questions about what they wanted to do next, and then showing them how they might accomplish it. He worried if he was “over-facilitating,” providing too much guidance in their project process. However, Sam felt that they needed a bit more guidance than other families because they were still struggling with some computer basics like using the mouse. Sometimes Sam had to put his hand over Eric’s hand on the mouse to help him. He shared his feelings about over-facilitation during the team debrief. Another facilitator, Alex, shared similar feelings, but because the young girl he was working with accidentally lost her project, he felt it was necessary to step in more directly. Alex sat with her as she tried to recreate her project.

During an extra workshop session before the community showcase, Sam shared an idea about adding “riders” and he modified a plastic cup with two googly eyes. Eric liked the idea and he and Julia each created their own. As he was watching them, Sam felt he had crossed into the over-facilitation and directed a part of their vision towards his ideas rather than theirs.

Case analysis
For Sam, it was a balancing act between his actions and ideas and Julia and Eric’s actions and ideas. This constant tension has been highlighted in other studies of facilitators (Gutwill et al, 2015). If a facilitator provides too much guidance and direction, learners may miss opportunities for discovery and exploration. On the other hand, too little guidance may frustrate and discourage learners as they run into challenges. In Sam’s experience, his practice of stepping in or out was informed by his close observation of Julia and Eric as well as critical reflection, on his own and with the facilitation team. Sam’s observations and interpretations of these interactions mirror the practices of researchers who collect data, make interpretations, and ask new questions. Eleanor Duckworth in The Having of Wonderful Ideas (1987) highlighted how the teacher is like a researcher:
Just as a researcher's knowledge guides her further questioning and gives rise to the next problem she asks them to consider, so a teacher, convinced that he cannot put his own understanding into the learners' heads, uses that understanding to help the learners take their own thoughts further. (Duckworth, p. 162)

In a post-workshop series interview, Sam continued to reflect on his facilitation between workshops. During Workshop 3, he felt like he “under-facilitated,” giving them too much space. He shared, “I wanted them to be proud of what they're doing and I wanted them to realize their vision, so if I maneuver them a little more effectively, I thought it would be a better overall experience for them.” His uses of the words “over” and “under” highlight the continual balancing act of facilitators. As he continued to help them, he struggled with that balance. Through critical reflection on his experience as well as with other facilitators, he later came to understand an important question for him to assess if he was over-facilitating: whose ideas are being represented? To him, over-facilitation occurred when his ideas were incorporated rather than their ideas.

Alex: Building relationships with learners

Alex, a software developer, joined the facilitation team late and started at Workshop 2. When he was in college, he facilitated technology-based camps for middle and high school students. However, he was new to this community site as well as to engaging families with limited backgrounds in computing.

At dinner during Workshop 2, Alex sat with another family who came late to the series: a retired great-grandmother, Sarah, who came with her 10-year old great-granddaughter, Mariah. Sarah shared with Alex how she was here for Mariah since Mariah’s mother could not attend. During Share, Sarah, who had limited mobility and moved with a walker, was unable to walk around each project. Facilitators including Alex noticed this challenge and discussed ways to better include and support Mariah in future workshops.

In Workshop 3, Alex noticed that Sarah came in with a cane. During dinner, Sarah shared with Alex that it was her first time not using her walker for 2.5 years because she felt comfortable that people would take care of her in this workshop. Her comment impressed him. He sat with them as they got started on their family project. Mariah sat forward as she worked on the laptop, while Sarah sat back watching Mariah and occasionally asking her questions. During team debrief, he shared Sarah’s decision to use a cane with the rest of the facilitators during the debrief. Jorge, another facilitator who worked as staff at the community center, praised the team for helping Sarah feel a sense of physical safety.

In the next workshop, Mariah came alone as there was a miscommunication with Sarah about the workshop. Alex watched as Mariah signed into her Scratch account, she was unable to find her last project. Mariah seemed upset and Alex became anxious. He tried to reassure her that they could recreate it. Alex wondered how much he should guide her and remembered one of the FCL facilitating fundamentals was to “ask questions rather than give answers.” He felt conflicted and checked in with another facilitator, who shared his anxiety. He decided to work closely with her so that she would not fall behind. As she shared ideas, Alex made suggestions for how they could implement it. Together they made a pool party with friends dancing to music by Beyoncé. Throughout the process, Alex made encouraging comments about her ideas.

Sarah attended the next workshop with Mariah. In addition to her cane, Sarah carried a large bag containing a blanket she knitted for Mariah’s family. She proudly showed it off during dinner and again during the Parent Meet session. When the Make portion of the workshop started, Mariah opened their Scratch project, while Sarah took out her knitting needles and started to work on another blanket. Alex noticed immediately and was worried that Sarah might be disengaged for the remainder of the workshop. However, he knew from his earlier conversation with Sarah that she came to the workshops for Mariah — something that Sarah repeated to him multiple times. He sat next to them and asked Mariah to explain her project progress to Sarah. Once Mariah began describing the project, Sarah put down her knitting work and started making suggestions.

Case analysis

In a post-workshop interview, Alex shared how proud he felt in supporting Sarah and Mariah, particularly how he re-engaged Sarah when she brought out her knitting needles. He intervened, but without being intrusive by asking Mariah to describe the project as a way to bring Sarah in. He also discussed the feeling of being part of a family. He shared, “Having to be a part and not a part another family’s dynamic [was new]. It’s really interesting to see how that works and to understand my role in that when it comes to both the social and the technological aspects of it.” Alex provided technical support when they needed it, but he also provided emotional and social support with his encouragement when they faced challenges.

In a study of families in a makerspace, Brahms and Crowley (2014) described the ways in which facilitators can become part of a family’s learning dynamic and an important social resource. Brahms and
Crowley noted that as children engaged in making, they are not only building relationships with materials, but also with the people that support them. For Sarah, her relationships with facilitators helped her feel comfortable. For Alex, relationship building was an instrumental part of his development as a facilitator as it helped him know how to intervene and interact with families. In his past experiences helping with technology-based camps, Alex shared how their top priority was making sure kids were continually “doing stuff and being engaged.” However, in these workshops, facilitators were encouraged to build relationships with families (which was one of the Facilitating Fundamentals). For example, through facilitators’ dinner conversations with families, he learned about Mariah’s interests and Sarah’s main motivation, which was being present for Mariah.

**Discussion**

This research highlights the different possibilities for adults to engage in facilitation. There was not one role that defined a facilitator, but instead multiple entry points and pathways that built on facilitators’ backgrounds and interests. Their experience was a process of becoming — shifting roles, perspectives, and practices within the shared activity of facilitating creative computing experiences (Lave & Wenger, 1992; Rogoff, 1994). Maria had a background as an educator with limited technology experience. Her role shifted to being a learner and her views of parents expanded to see them as joyful and curious learners. While Sam had some experiences teaching with technology, he continued to negotiate his level guidance and he learned what mattered: families’ ideas. Alex, as a software developer, had technical experience to contribute, but learned the value of building relationships and connecting with families in his role as a facilitator.

We are also interested in what supported these adults to take on the responsibilities and practices of facilitation. First, the facilitation team became an important context for facilitators’ development as they learned from one another in the shared activity of supporting families’ learning (Lave & Wenger, 1991; Rogoff, 1994). Facilitators recognized and pulled on one another’s strengths and complemented their varying backgrounds. For example, Maria connected with Alex to troubleshoot families’ challenges with Scratch. Secondly, collective and critical reflection allowed facilitators to strategize around challenges as well as provide emotional support as facilitators confronted their different anxieties. Sam and Alex’s discussions about over-facilitation helped them understand how to recognize if their actions crossed the line into over-facilitation or if their actions still aligned with their values of helping learners realize their ideas. Maria shared her insecurities around her technical expertise during team debrief and received feedback from other facilitators on how to shift her practice towards a tinkering mindset. This collective and critical reflection was supported by their personal field notes as well as team meetings between and after workshops, where they discussed close observations, asked questions, and provided suggestions. Finally, building relationships with families was important for facilitators to understand families’ needs, interests, and motivations. This understanding informed their decisions as facilitators. For facilitators, eating dinner and connecting with families was just as important as working together on projects.

In these coding and making experiences with technology, staff at youth programs or other informal learning spaces might lean towards recruiting students, professionals, and other volunteers who have expertise with computing. However, these facilitators’ trajectories highlight the variety of expertise needed to support families in a creative learning experience. Maria and Sam had backgrounds as former teachers and children with special needs. The community staff played important roles in helping facilitators understand the backgrounds of the different families. Additionally, these facilitators’ experiences highlight what practices might be undervalued when we focus on STEM knowledge and expertise. For example, Maria felt anxiety about her limited experience with programming. However, facilitators highlighted the importance of modeling exploration and experimentation as well as providing encouragement and suggestions. These facilitation moves have also been highlighted in other studies of facilitators (Gutwill et al., 2015; Vossoughi et al., 2013).

This study is a first step in deepening our understanding of what supports facilitators to take on the many roles that they play to facilitate creative and inclusive learning environments in computing. We are interested to continue exploring what might support adults to carry out different facilitation practices, such as surfacing learners’ interests or encouraging exploration and experimentation while facing different challenges. For example, we hope to explore some of the tensions that emerged in facilitators’ experiences. How can facilitators help guide learners’ exploration, development, and implementation of their ideas, while under the constraints of a workshop’s time limits or other obstacles (e.g. lost projects or challenging collaborative dynamics)? The FCL environment created an interesting context to study the dynamics between facilitators and families and what facilitation strategies matter when supporting a family unit, each with its own dynamics and tensions between parents and their children. Future research could contribute to how facilitators support family dynamics in the context of creative computing experiences where different family members have different expertise and perspectives on computing. For some families, the role of an expert parent and novice child are reversed in the context of computing as children develop their expertise and confidence beyond their parents’
abilities (Correa, Straubhaar, Chen, and Spence, 2013). Additionally, we understand that not all informal learning contexts with creative computing have the level of engagement and structure like FCL, but we believe that calling out the different features of facilitators’ experiences in FCL could be helpful to other settings, such as drop-in sessions in some museums and other learning environments (Brahms and Crowley, 2016). More research could focus on the development of facilitators in settings beyond workshops.

Often when out-of-school organizations onboard facilitators, the mindset is to “train” facilitators. However, our studies of facilitators’ experiences and the roles that they enacted suggest, instead, to “facilitate the facilitators.” We argue that learning environments should support a community of learning where facilitators also take on the role of learners — people who are learning to welcome, support, and guide diverse learners in designing, tinkering, and making with computing. In these case studies, facilitators were given many opportunities to support one another, to reflect and discuss their experiences, and to build relationships with workshop participants. These were valuable opportunities to allow them to become a facilitator. In her post-workshop interview, Maria wondered “Am I the right fit?” when she was recruited as a facilitator. She later learned through her experience that fitting in was not necessary because FCL was learning environment. “I became the right fit, and FCL became the right fit for me. It's a learning experience.”

References
A Climate of Support: A Process-Oriented Analysis of the Impact of 
Rapport on Peer Tutoring

Michael Madaio, Human-Computer Interaction Institute, Carnegie Mellon University, mmadaio@cs.cmu.edu
Kun Peng, Human-Computer Interaction Institute, Carnegie Mellon University, aeo@cs.cmu.edu
Amy Ogan, Human-Computer Interaction Institute, Carnegie Mellon University, justine@cs.cmu.edu
Justine Cassell, Language Technologies Institute, Carnegie Mellon University, kunp@andrew.cmu.edu

Abstract: Prior work has found benefits of interpersonal closeness, or rapport, on student learning, but has primarily investigated its impact on learning outcomes, not learning processes. Moreover, such work often analyzes the direct impact of dyadic features like rapport on learning, without considering the role played by individual factors, such as learners’ prior knowledge and self-efficacy. In this paper, we investigate the intertwined impact that rapport, self-efficacy, and prior knowledge have on the process and outcomes of peer tutoring. We find that peer tutors in high-rapport dyads offer more help and prompt their tutees to explain their reasoning more than low-rapport dyads, with tutees in such dyads verbalizing their problem-solving process and proposing more steps and answers. Meanwhile, rapport is associated with increased procedural performance, but tutees’ self-efficacy and prior knowledge moderate the effect of rapport on tutees’ conceptual performance.

Introduction
As decades of scholars have argued, learning is a socially-mediated process, embedded in relationships between students and their teacher, and between students and their peers. The interpersonal components of classroom talk - those focused on building and maintaining social relationships - are inextricably intertwined with the interactional process of providing feedback, hints, and instructions (Wentzel, 1999; Parr and Townsend, 2002; Madaio et al., 2017). Though teachers and schools may primarily promote educational goals, students arrive in the classroom with relational goals, which may include developing relationships and friendships with each other (Wentzel, 1999). They also arrive with individual resources for and attitudes towards learning that may impact the ways that they establish and respond to both the educational and social environment of learning. Prior research has particularly targeted reciprocal peer tutoring as a form of learning likely to result in learning gains due to relational factors (Parr and Townsend, 1992; Webb and Mastergeorge, 2003). As students are asked to learn in increasingly collaborative learning environments, and as one of the most touted 21st century job skills is the ability to collaborate, it is important to understand whether and how the rapport learners build with one another impacts learning. Specifically, there remain questions about how the social bond among students intersects with the individual resources and beliefs of each student to impact the process and outcomes of peer tutoring.

Educational technologies are increasingly designed to mediate students’ collaborative learning (Carmien et al., 2007), or to take on the role of a teacher or peer. In prior work, we have treated the question of how the social relationships between peers in collaborative learning (here, peer tutoring) impact learning gains. Here we ask how rapport's impact on learning gains might be mediated by the student's own academic ability and beliefs about learning (which we operationalize here as self-efficacy; Pajares, 1996). We have argued that rapport is a phenomenon that must be understood – and designed for – as taking place over time (Zhao et al., 2014). For that reason, in this paper we look not just at learning gains, but also at the process-oriented mechanisms by which social factors interact with individual cognitive and socio-cognitive factors to influence the peer tutoring process. Specifically, in this work we investigate how rapport between peer tutor and tutee influences the learning-related behaviors that each participant engages in, moderated by tutees’ prior knowledge and self-efficacy. We contribute to the larger learning science community by (1) Shedding light on how the impact of social factors like rapport on learning is moderated by individual cognitive and socio-cognitive factors like prior knowledge and self-efficacy, respectively; (2) Identifying the core sequences of tutor and tutee behaviors associated with moments of high and low rapport, to shed light on the peer tutoring process as it plays out over time; and (3) Teasing apart the intertwined influences of rapport, self-efficacy, and prior knowledge on tutors’ help-provision and tutees’ verbalization of problem-solving.

Related work
Rapport has been described as a “harmonious, coordinated interaction based in mutual attention, respect, understanding, and openness”, characterized in learning interactions by nurturance on the part of a teacher and
increased engagement in classroom activities on the part of students (Wentzel et al., 1999). However, in contrast to prior work on teacher-student rapport, the impact of student-student rapport on peer learning interactions has been less well-studied. Work in our own lab suggests that rapport is beneficial for learning outcomes in peer tutoring settings (Zhao et al., 2014; Sinha and Cassell, 2015), but that work did not look at whether or how individual factors mediated the impact of rapport on learning processes. Work on differences in learning dynamics between dyads of friends and non-friends (often a proxy for rapport) found that learning interactions between friends are characterized by a warm, supportive climate favorable to exploration and problem-solving, a more elaborated learning discourse, and greater expectations for assistance and support from the other, (Wentzel, 1999), as well as greater tolerance for direct challenges and differences of opinion (Azmitia and Montgomery, 1993). Prior work suggests that the supportive climate engendered by relational closeness may constitute a motivational support that mitigates anxieties around self-explanations, help-seeking, and problem-solving attempts, encouraging more participation and engagement in learning (Wentzel, 1999). However, these findings were based on groups of self-defined friends, which may mask a great deal of variation in the nature or duration of those friendships. To unpack these nuances, we focus on students who have never met before.

From the tutor’s perspective, prior work suggests that more capable peer tutors support their partners’ needs for “positive face”, or desire to be approved of by others (Brown and Levinson, 1987), particularly when they lack the relational closeness of friendship or rapport (Madaio et al., 2017). Peer tutors may do this by modifying the behaviors that may threaten the other’s face, such as overly direct instructions or knowledge monitoring, while increasing “face-boosting” behaviors, such as positive feedback. Other work has argued that close, “integrative” social relationships among peers are characterized by greater resource provision (Wentzel 1999) and unsolicited advice- or help-giving, which may be perceived as face-threatening without a sufficient relationship (Feng and Magen, 2013). However, prior work suggests that tutors’ provision of the correct step or answer to a problem without providing deeper conceptual support may be not only unhelpful, but may actually be harmful for tutees’ learning (Webb and Mastergeorge, 2003). The best tutors, therefore, prompt their students to provide elaborated explanations of the concepts and reasoning either through explicitly requesting explanations or by prompting them with leading questions (Webb and Mastergeorge, 2003).

While this work demonstrates that relational closeness has a positive influence on students’ participation and engagement in learning, other prior work has uncovered competing influences at work: namely, students’ self-efficacy, or their beliefs in their ability to successfully complete a task, and their prior knowledge (Webb and Mastergeorge, 2003; Wentzel, 2016). Students with lower self-efficacy or prior knowledge may be more likely to avoid publicly engaging in the problem-solving process by verbalizing fewer self-explanations and attempts of the problems, due to fears of judgment or negative evaluations from their peers (Wentzel, 1999), which may conflict with the motivating influence of rapport. Such students might not only be reluctant to attempt the problems for fear of failure but, worse, they might engage in what Newman (2002) describes as “nonadaptive” or “dependency-oriented” help-seeking, or, help requests designed to elicit the answer from a partner to avoid the risk of failing. Therefore, it is critical to understand how the help provided by peer tutors is influenced by their rapport with their tutees, and how tutees’ responses to tutors’ help are influenced by their rapport with their partner, their self-efficacy and their prior knowledge.

In sum, prior work argues that interpersonal closeness between tutor and tutee is likely to lead to increased tutor helping behaviors and increased tutee engagement in the learning process. However the nature of that interpersonal closeness has primarily been treated as a dichotomous measure of friendship, which may mask variation in the nature and strength of those friendships. We thus investigate the rapport established between strangers, to identify how peers without an existing relationship negotiate between their relational and their task goals. Further, prior work suggests that learners’ self-efficacy and prior knowledge may moderate the pathway from social support to improved learning outcomes (Wentzel, 2016), but it leaves open the question of how rapport, self-efficacy, and prior knowledge interact to impact tutoring processes and learning outcomes. This leads us to investigate the following research questions:

**RQ1**: To what extent does the rapport between peer tutors and their tutees, when they first begin their relationship, interact with those tutees’ prior knowledge and self-efficacy to impact their learning outcomes?

**RQ2**: How does the rapport between peer tutoring dyads impact the learning process, as it plays out over time?

**RQ3**: To what extent does the rapport between newly formed tutoring dyads interact with tutees’ prior knowledge and self-efficacy to impact tutees’ help-seeking and self-explanations, and tutors’ help provision?
Methodology

Data corpus and dialogue annotations
The corpus described here was collected as part of a larger research program on the effects of rapport-building on reciprocal peer tutoring. Participants were assigned to 20 dyads that alternated tutoring one another in linear algebra equation solving for two hour-long sessions, a week apart. Though data from only 15 dyads were usable due to issues with recording, participant retention, or a pre-existing relationship between partners, for a total of ~30 hours of data. Each session was structured such that students engaged in two brief periods of getting to know one another interleaved with two longer periods of tutoring, where each student was randomly assigned to be the tutor for one of the tutoring periods in each session. The peer tutor was not assumed to have any greater prior knowledge than their partner for the problems they were tutoring them on and, as such, all tutors were provided with instructions on how to teach the problems for which they were assigned the role of tutor.

The students took a pre-test before the first session and a post-test after the final session to assess their prior knowledge and their learning gains. Both tests included procedural (problem-solving) and conceptual items (multiple choice items on algebraic concepts). The participants (mean age = 14.3, min=13, max=16) came to a lab on an American university campus in a mid-sized city for the study. Half were male and half female, assigned to same-gender dyads, so that, as in other work with this corpus, any gendered differences in the social, rapport-building behaviors of the participants could be identified. To control the baseline level of interpersonal closeness, we paired students with a partner they had not met before (and removed two dyads that we later discovered to have already met), similar to the dyads of “strangers” in Madaio et al., (2017). Audio and video were recorded, and audio transcribed and segmented into clauses. Participants were provided with survey items to assess their self-efficacy for algebra, following (Pajares et al., 1996), and their perceptions of the rapport in the dyad. As part of a larger research program on the relationship between rapport and peer tutoring, this corpus was annotated for a set of social, rapport-building verbal and nonverbal behaviors, following Zhao et al., (2014), and for a set of tutoring and learning behaviors, following Madaio et al. (2016)’s work on knowledge-telling, knowledge-building, procedural and conceptual questions (though here we additionally annotated for subtypes of each of those categories, as shown in Table 1, with Krippendorff’s alpha for all categories > 0.75). In this paper, we describe analyses of behaviors likely to be impacted by relational closeness: two types of tutor behaviors: tutors’ help-offering and explanation-prompts; and five types of tutee behaviors: tutees’ help requests, answer-checking, step-level procedures, answers-proposed; and evidence-based reasoning.

Table 1: Tutoring and Learning Strategies Annotated

<table>
<thead>
<tr>
<th>Role</th>
<th>Low-Level Strategy</th>
<th>Abbreviation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutor</td>
<td>Help-Offering</td>
<td>HO</td>
<td>“Let me know if you need any help with this one”</td>
</tr>
<tr>
<td>Tutor</td>
<td>Explanation Prompting</td>
<td>EP</td>
<td>“Why did you say to divide it by 5?”</td>
</tr>
<tr>
<td>Both</td>
<td>State Step-Level Procedures</td>
<td>SP</td>
<td>“Ok so I’m gonna add three to both sides here.”</td>
</tr>
<tr>
<td>Student</td>
<td>General Help Requests</td>
<td>HR</td>
<td>“Wait, how do I do this kind again?”</td>
</tr>
<tr>
<td>Student</td>
<td>Answer Checking</td>
<td>AC</td>
<td>“Is that right?”</td>
</tr>
<tr>
<td>Student</td>
<td>Answer Proposing</td>
<td>AP</td>
<td>“I think x is 7.”</td>
</tr>
<tr>
<td>Student</td>
<td>Explain Reasoning</td>
<td>ER</td>
<td>“I divided by 5 because it needs to be reduced.”</td>
</tr>
</tbody>
</table>

Rapport rating
Rapport between dyad participants was evaluated using a “thin-slice” approach (Ambady and Rosenthal, 1992), where the corpus was divided into 30-second video slices provided to naive, third-party raters, following Sinha & Cassell (2015) where three raters rated the rapport in each slice on a Likert scale from 1-7, and a single rating was chosen for each slice using a bias-corrected weighted majority vote approach. From the 120 30-second slices in each session, we calculated a summary rapport score for each session, the “utopy” (Sinha, 2016), which we use in this paper as our rapport measure. Prior work has shown that statistical summaries such as a measure of central tendency or proportion of high and low ratings of rapport collapse the temporal dimension and are not as robust as more stochastic-based models which capture the evolution of rapport over time (Sinha, 2016). We thus fit a Markov chain of order 1 to the sequence of 120 rapport ratings for each session, and used the resulting transition probability matrix to generate a measure of the “utopy”, or likelihood of the dyad being in a high-rapport state, which is the sum of each transition probability weighted by the distance of the transition (e.g. rapport 2 to 5) (Madaio et al., 2017; Sinha, 2016).

Sequence mining and clustering
To better understand how the rapport between tutor and tutee impacts the tutoring and learning process over time, we adopt a sequence mining approach. We used the USpan algorithm to mine “high-utility” sequential patterns, where “utility” in this case is the rapport value of the thin-slice window in which a given behavioral sequence occurred (Fournier-Viger et al., 2016). However, most sequential pattern mining algorithms produce large outputs with many similar patterns that make interpretation of the resulting sequences difficult. To resolve this, we developed a sequence summarization approach, based on Patel et al.’s sequence clustering technique (Patel et al., 2017). A directed weighted graph was constructed to represent a set of sequences as a cluster, with each annotated behavior represented by a node, and the number of times each multiple-node behavior sequence occurred represented by the weights of the edges, similar to (Patel et al., 2017). We differ from Patel et al. by then summarizing each sequence cluster into a single “core sequence” which reflects the most representative behavior sequences for that cluster. For each graph, we first simplify the result by removing unrepresentative edges (those with very small weights). We then identify the “core sequence” of each sequence cluster, defined as the longest simple path between any two nodes on the graph to find the most common sequences.

**Findings**

We first wanted to validate our use of the “utopy” measure of rapport, or, the likelihood that rapport will be increasing from one “slice” to the next, by understanding its relationship to students’ self-reported rapport at the end of each session. Prior work has found a significant correlation between “thin-slice” observer judgments of interpersonal rapport and participants’ own self-report (Grahe and Bernieri, 1999). Our data replicates that finding, showing a significant positive correlation between the “utopy” measure of rapport and that dyad’s average self-report rapport rating ($p = 0.49$, $p<.005$). Other summary measures of the thin-slice rapport data, such as the average value for each dyad, did not, following Sinha’s (2016) finding that utopy was a more informative measure of temporally dynamic interpersonal constructs such as rapport than simple measures of central tendency such as the average. For the following analyses, all results are significant after correcting for multiple hypothesis testing using the Benjamini-Hochberg post-hoc correction (Benjamini and Hochberg, 1995).

**Rapport and learning outcomes**

We first investigated research question (RQ1): how tutor-tutee rapport interacts with tutees’ self-efficacy and prior knowledge to predict tutees’ learning outcomes. Since prior work suggests that students with lower prior knowledge and self-efficacy will reap the most benefits from the socially supportive environment engendered by rapport (Wentzel, 2016), we hypothesized that students with lower prior knowledge and self-efficacy will demonstrate greater performance at post-test when they have increasing rapport with their tutors (i.e. high utopy). In order to identify interactions between tutees’ prior knowledge (operationalized by their score on a pre-test) and their rapport, we used the post-test performance as the dependent variable rather than a normalized learning gain score. We ran two hierarchical linear mixed-effects models, with two measures of tutees’ post-test performance (procedural items and conceptual items) as the respective dependent variables, and interactions between tutees’ pre-test percent, their self-reported self-efficacy, and their rapport utopy measure as fixed effects, with random effects for session and dyad. Significant results are reported in Table 2.

**Table 2: Coefficients of regressions for pre-test, rapport, and self-efficacy on procedural and conceptual items**

<table>
<thead>
<tr>
<th>Post-test Item Type</th>
<th>Pre-test</th>
<th>Self-Efficacy</th>
<th>Rapport</th>
<th>Rapport* Pre-test</th>
<th>Rapport* Self-efficacy</th>
<th>Rapport*Pre-test Self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedural</td>
<td>0.47***</td>
<td>0.28***</td>
<td>0.40***</td>
<td>0.10*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual</td>
<td>0.67***</td>
<td></td>
<td>-0.25 **</td>
<td>0.19 **</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the model with the procedural post-test items only, rapport was only predictive while interacting with prior knowledge and with self-efficacy. In those two two-way interactions, while tutees in high-rapport dyads did demonstrate greater procedural performance over tutees in low-rapport dyads, unexpectedly, the tutees with high-rapport with their tutors that also had greater prior knowledge and greater self-efficacy performed better than their low prior knowledge and low self-efficacy peers on the post-test items. That is, contrary to our hypothesis, tutees that already came to the session with greater prior knowledge and greater self-efficacy benefited the most on procedural test items from having high rapport with their tutors.
The conceptual post-test performance tells a different story, however. There was a two-way interaction between rapport and self-efficacy, and a three-way interaction between rapport, self-efficacy, and prior knowledge. In the three-way interaction, the tutees with low prior-knowledge and low self-efficacy demonstrated greater conceptual performance when they had high rapport with their tutor, in line with our hypothesis that students who needed the most support would benefit from rapport (Wentzel, 2016). Unexpectedly, however, tutees with low prior-knowledge and high self-efficacy demonstrated worse conceptual performance when they had high rapport with their tutor. This same pattern was true for the two-way interaction between rapport and self-efficacy as well. That is, when tutees with lower self-efficacy had increasing rapport with their tutor, they were more likely to have higher conceptual post-test scores, but when tutees with greater self-efficacy had higher rapport with their tutor, their conceptual post-test scores were more likely to be lower.

Unfolding processes over time: Core sequences for high and low rapport
To understand how the rapport between tutor and tutee impacted their respective contributions to the intertwined, dyadic process of tutoring and learning (not just the learning outcomes on the post-test), we conducted a sequence mining and clustering analysis to identify the most common sequences of tutoring and learning behaviors. We will discuss here the most commonly occurring “core sequences” of dyadic behaviors associated with high-rapport and low-rapport values in the 30-second “thin-slice” windows (RQ2).

The core behavioral sequence cluster associated with high rapport (rapport values above 4), which occurred a total of 151 times in our corpus (out of a possible 834 sequences), contained the behavioral sequence {Tutee States Procedures, Tutee Explains Reasoning, Tutor Help-Offering}. That is, the tutee first verbalized some step-level procedure, then explained their reasoning for doing so, and only then did the tutor offer to help. Such a process is characterized by self-explanations from the tutee and the subsequent provision of help offered by the tutor, as in Wentzel (1999). One example of this from our corpus is the tutee saying, “Wait, so it’s k plus k plus two” (SP), then “Cause I need to get all the k’s on one side” (ER), followed by the tutor, “Does it make sense what you need to do after that?” (HO). It is significant that we see this order to the sequence, rather than the opposite: tutors offering help prior to the tutee attempting the problem, (which was not one of the core sequences associated with high rapport), as that sort of anticipatory help-offering may imply a lack of confidence on the part of the tutor for the tutees’ ability to solve the problem on their own, rather than offering support after the tutee attempts to solve it first (Webb and Mastergeorge, 2003).

The core behavior sequence associated with low rapport (rapport values below 4), which occurred 268 times in our corpus, contained the behavior sequence {Tutee Proposes Answer, Tutee Checks Answer, Tutor States Procedure, Tutee States Procedure}. That is, we see a problem-solving process characterized by tutees proposing an answer, requesting confirmation of the correctness of that answer, and then the tutor providing the next step, which is then stated by the tutee. In our corpus, one instance of this is the tutee saying, “y is equal to negative five” (AP), “Is that right?” (AC), and the tutor responds with “Try adding the 10 to both sides instead.” (SP), followed by the tutee; “Ok, so ten plus five is fifteen.” (SP). Here, the tutor is explicitly requesting the help or feedback that may instead be provided by the tutor unsolicited if they had greater rapport between them. This also suggests the type of nonadaptive, dependency-oriented help-seeking described by Webb and Mastergeorge (2003), where the tutee requests feedback and validation throughout the problem-solving process that they may not need, rather than trusting in their ability to solve the problems or in their tutor to provide help.

Impact of rapport on tutoring and learning behaviors
Given the surprising impact on learning outcomes from the interaction between rapport and individual tutee factors (prior knowledge and self-efficacy) in RQ1, we wanted to understand how those factors jointly impacted each individual tutee and tutor behavior, to complement our sequential analysis of those intertwined behaviors (RQ2). Specifically, we wanted to understand how dyadic rapport interacts with tutees’ self-efficacy and prior knowledge to influence tutees’ learning strategies (i.e. self-explanations and help-seeking behaviors) and tutors’ instructional behaviors (i.e. help-offering and explanation-prompting). We hypothesized (RQ3a) that tutors will do more explanation-prompting and help-offering when rapport is increasing, based on prior work that suggests that friendship, or a social relationship, allows for the kinds of probing questions (Azmitia and Montgomery, 1993) and unsolicited advice-giving (Feng and Magen, 2013) that might be seen as face-threatening without that relationship. We also hypothesized (RQ3b) that tutees with lower prior-knowledge and lower self-efficacy will generally request help and check their answers more often in dyads with high rapport, based on prior work that suggests that students’ willingness to seek help may be inhibited without a sufficient social relationship with peers (Newman, 2002). We also hypothesized that tutees in high-rapport dyads will verbalize their problem-solving more readily than in low-rapport dyads, particularly those with greater self-efficacy, following prior work suggesting that friends (a proxy for high-rapport) are more likely to take risks in learning due to a reduced
fear of failure (Azmitia and Montgomery, 1993), and prior work in peer tutoring that found that tutees who were friends with their tutors engaged in more “knowledge-telling” (Madaio et al., 2016).

To investigate these hypotheses, we ran a set of hierarchical linear models with the frequency of tutor and tutee behaviors of interest in each session as the dependent variables, normalized by their total utterances, and with tutors’ and tutees’ pre-test percent, their self-efficacy, and the rapport utopy score for that session as the fixed effects, with interaction terms, and with random effects for session and dyad. We ran these models for tutors’ help-offering (HO) and explanation prompting (EP), as well as tutees’ stated procedures (SP), answer-proposing (AP), generic help-seeking (HS), answer-checking (AC), and explanations of reasoning (ER) as the DVs for each model. We report significant results from the models in Table 3.

Table 3: Coefficients of regressions for pre-test, rapport, and self-efficacy on tutoring and learning strategies

<table>
<thead>
<tr>
<th>Tutoring and Learning Strategies</th>
<th>Tutees’ Pre-test</th>
<th>Self-Efficacy</th>
<th>Rapport</th>
<th>Rapport* Pre-test</th>
<th>Pre-test* Self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutor Help Offering</td>
<td>-0.43***</td>
<td>0.33*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explanation Prompting</td>
<td>0.37*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tutee General Help Requests</td>
<td>-0.31*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Answer Checking</td>
<td>-0.38**</td>
<td>-0.36**</td>
<td>-0.23*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Answer Proposing</td>
<td>0.43**</td>
<td>0.45**</td>
<td>-0.56***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stating Procedures</td>
<td>0.40*</td>
<td>-0.40**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explain Reasoning</td>
<td>-0.30*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the tutors’ help-offering and explanation-prompting, as we hypothesized, when in sessions with rising rapport (utopy), tutors prompt their tutees for explanations of their reasoning more often (e.g. “wait, why would it be two?”) and offer more help to their tutees (e.g. “do you want to see my work?”). This aligns with results from the high-rapport sequences mined. Tutors also offer help more often when tutors’ prior knowledge is low, suggesting that peer tutors not only recognize when help is needed (i.e. when their tutees have low prior knowledge), but they are also influenced in their help provision by the rapport they have with their tutee.

For tutees, low prior-knowledge tutees use general help requests (e.g. “wait where do I even start?”) more often than high prior-knowledge, as was hypothesized, with no effect of rapport or self-efficacy. Similarly, for tutees’ answer-checking (e.g. “is it nineteen over six?”), tutees with lower prior knowledge check answers significantly more than those with greater prior knowledge. Contrary to our hypothesis about rapport motivating more help-seeking, tutees with higher rapport ask their tutors to check their answers significantly less than those with lower rapport with their tutors, which was also seen in the low-rapport sequences mined in RQ2. It may be the case that, given the supportive environment of high rapport, these tutees did not feel the need to explicitly request the help that was freely offered in those high-rapport dyads. Surprisingly, tutes with low prior-knowledge and higher self-efficacy answer-check significantly more than tutes with both low prior-knowledge and low self-efficacy. This increased use of explicit verification of answers suggests a dependency-oriented approach to help-seeking exhibited by low prior knowledge tutes, particularly when those tutes also have high self-efficacy, aligning with our findings from RQ1. This also aligns with Williams and Takaku’s (2011) finding that middle-school students with high self-efficacy, despite their low performance, engaged mostly in non-adaptive help-seeking, which included seeking help when not needed.

For tutes’ answers proposed (e.g. “p equals 23”), tutes in high-rapport dyads are proposing more answers than those in low-rapport dyads, which may be explained by the reduced apprehension of critical feedback in a socially-supportive environment engendered by rapport. Interestingly, low prior knowledge students with high self-efficacy proposed more answers than all others. That is, these students who have confidence in their algebra ability, despite their low performance on the pre-test are both proposing more answers, and asking for confirmation of those answers (as seen in the previous result). For the tutes’ stated procedures (e.g. “I think I should add thirteen”), high rapport is associated with more verbalizing of their problem-solving procedures, as hypothesized. For the low prior knowledge tutes, those in high rapport dyads verbalize their problem-solving procedures significantly more than those with low rapport. This aligns with our
hypothesis that students with less prior knowledge who might otherwise feel apprehensive about verbalizing their problem-solving process are more likely to do so when they have greater rapport with their tutor. Similarly, for the explanations of reasoning (e.g. “I’m gonna divide cause I’m trying to cancel out”), low prior knowledge tutees in high-rapport dyads explain their reasoning more than those with low-rapport.

These results suggest that rapport contributes to a supportive climate for learning where help from the tutor is provided more readily and tutees feel encouraged to verbalize their problem-solving process. However, for students whose confidence in their abilities outstrips their actual performance, the increased verbalization of the problem-solving process may contribute to their use of a more dependency-oriented, nonadaptive help-seeking process. While this may lead to improved procedural performance, without engaging in elaborated explanatory discourse, this may cause them to miss opportunities to improve their conceptual understanding.

Discussion
Designers of learning environments, technological or otherwise, are always designing—whether they are aware of it or not—for learners who are imbricated in webs of socio-cognitive relations with other learners, thinking and learning in partnership with other learners and with the technologies they use. Therefore, if designers of learning environments are attempting to intervene to improve the learning process, they must first understand the nature of the existing socio-cognitive processes at work. Carmien et al., (2007) have argued that the “scripts” introduced by collaborative support tools may conflict with the internal “scripts” that students themselves already bring to bear on the collaborative learning process. Prior work suggests that the rapport between teachers and students or between peers benefits learners by creating a warm, socially-supportive learning environment (Wentzel, 1999), while other work suggests that social factors in the interaction impact learning (e.g. Schnaubert and Bodemer, 2016). Until now, however, we have not had a well-defined, process-oriented understanding of how dyadic rapport intersects with individual features like prior knowledge and self-efficacy to impact the process of learning in peer tutoring.

In this paper, we identify a surprising interaction between rapport, prior knowledge, and self-efficacy on learning outcomes, and highlight the influence of each of those factors on tutoring and learning behaviors to shed light on the tutoring process. We find that tutors with greater rapport with their tutees offer more help and prompt for explanations more often, perhaps due to feelings of increased responsibility for their partner’s performance (Wentzel, 1999) or to increased freedom to engage in unsolicited help-offering, knowing that it is less likely to be taken as face-threatening (Feng and Magen, 2013). Given these results, it seems the peer tutoring process in dyads with high rapport is characterized by investment of the tutor in the comprehension and problem-solving process of their tutee, wherein learners are motivated by a socially supportive environment to self-explain, without needing to continually request confirmation of their answers. To understand why tutees whose self-efficacy outstrips their prior knowledge might rely more on an answer-proposing and –checking approach, we turn to Williams and Takaku’s (2011) finding that such students demonstrated a performance-oriented approach to learning, instead of a mastery-orientation (Williams and Takaku, 2011). For such students, perhaps obtaining the correct answers to the problems is more critical than correctly understanding the concepts, which may support our unexpected conceptual learning result. Webb and Mastergeorge (2003) also argued that dependency-oriented help requests may inadvertently signal to tutors that learners are unmotivated or unable to complete the problem, thus eliciting the steps or answers from the tutor, rather than the deeper conceptual explanations learners may need (Parr and Townsend, 2002).

It is in light of such results that designers of computer-supported collaborative learning (CSCL) tools, intelligent tutoring systems (ITS), or other artificial intelligence in education (AIED) applications should design systems that can help establish this supportive climate of mutual nurturance and support. One example is “socially-aware” virtual tutors or teachable agents that can build rapport with their partners (as described in Walker and Ogan, 2016). Such systems should thus incorporate adaptive socially-aware behavior patterns that elicit feelings of increasing rapport in the human learner, and be ready to deal with tutees’ learning behaviors increasingly in line with the findings here (i.e. verbalizing problem-solving steps, explaining reasoning, and proposing more answers).

This study is not without limitations as the small sample size limits the power of the findings. The findings here may also not be generalizable beyond the age group or culture of the participants represented. In addition, these analyses were correlational, as we did not experimentally manipulate the rapport of the dyad. Future work may first explore a causal inference approach such as Granger causality to infer the causal relationship between rapport and learning behaviors. We are also developing a rapport-building virtual peer tutor to attempt to build rapport with students in an experimental study, to identify causal relationships between rapport and learning. Conversely, a teachable agent might strategically build rapport with a student to elicit the supportive helping behaviors seen here. In sum, we intend for this paper to contribute to the growing body of
work understanding the socio-cognitive processes by which learning occurs. In this paper, we offer an analysis of how dyadic rapport built between students meeting for the first time interacts with individuals’ self-efficacy and prior knowledge to impact their learning outcomes and their tutoring and learning behaviors, and we offer one approach to unpacking the sequential process of tutoring and learning as it unfolds over time.

Acknowledgments

The research reported here was supported by the National Science Foundation Cyberlearning Award No.1523162, and the Institute of Education Sciences, U.S. Department of Education, through Grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. We would also like to thank our intrepid lab members for all of their hard work and stimulating conversations throughout the course of this work.

References

Augmenting Qualitative Analyses of Collaborative Learning Groups Through Multi-Modal Sensing

Bin Xie, Joseph Reilly, Yong Li Dich, and Bertrand Schneider
bix175@student.hks.harvard.edu, josephreilly@g.harvard.edu, ydich@college.harvard.edu,
ertrand_schneider@gse.harvard.edu
Harvard University

Abstract: In a previous study (N=84), we collected information about dyads who worked on an engineering task typical of makerspaces: programming a robot to solve mazes of increasing difficulty. We collected multimodal data using a variety of sensors, including mobile eye-trackers, galvanic skin response, motion sensors and audio/video streams. In this paper, we contrast two pairs that exhibited positive and negative learning gains. We first detail multimodal measures to compare differences and similarities across those groups, and then dive deeper into a qualitative analysis of their exchanges. We then describe how those measures could be used over the entire sample to capture productive interactions in small groups. We conclude by discussing how process data from sensors can augment traditional qualitative observations, and how it can create powerful synergies for better understanding collaborative interactions among learners in settings such as makerspaces.

Introduction

Supporting STEM learning (Science, Technology, Engineering, Mathematics) has become a primary focus of the Learning Sciences over the past decade. There is also a growing interest to understand how we can teach 21st century skills within those domains (e.g., Collaboration, Communication, Critical thinking, Creativity). The combination of those factors has contributed to the popularity of makerspaces. Makerspaces are informal learning environments where students learn complex concepts in STEM by building their own artifacts using digital fabrication tools (e.g., laser cutters, 3D printers, robotics). We are interested in understanding what promotes learning in those spaces - especially from a socio-constructivist perspective (Palincsar, 1998). We presuppose that social interactions are among the main drivers of learning, because students spend a significant amount of time interacting with their peers and facilitators. This paper is about conducting a multimodal analysis of a typical makerspace activity, and isolating factors that contribute to productive collaborations. More specifically, we isolated two pairs from a larger study (Starr, Reilly & Schneider, 2018) and are qualitatively analyzing their interactions. The main contribution of this paper is that we are leveraging methods from the field of Multi-Modal Learning Analytics (MMLA; Blikstein & Worsley, 2016) to support our qualitative observations and helps us generate measures of productive interactions in small groups.

The paper is structured as follows: first, we conduct a literature review of indicators of collaboration from a multimodal perspective (physical, physiological and visual synchronization). Second, we summarize the study and present the two pairs that we are contrasting. Third, we analyze those two groups using a variety of qualitative and quantitative methods. We leverage sensor data to augment our two case studies using data from eye-trackers, a motion sensor and wristbands capturing electrodermal activity. Fourth, based on those analyses we present measures of synchrony that we plan to extend to our entire sample of 42 pairs. We conclude by summarizing our results and discussing next steps.

Literature review

As a first step and for the scope of this paper, we are focusing on measures of synchronization in small groups. We review three kinds of synchronization that could characterize productive interactions: physical, visual and physiological. Because of space limitations, we only discuss the main contributions of each domain.

Physical synchronization

The synchrony of physical movements within groups using multimodal learning analytics is an emerging aspect of research on collaboration. Behavioral coordination between group members is generally indicative of positive outcomes and has been studied extensively (Pentland & Heibeck, 2008.) This type of qualitative analysis, however, is time-consuming and requires expert knowledge of gestures to code correctly. Using sensor data and computational methods, Worsley and Blikstein (2013) have pioneered new ways to study embodied learning and found that experts in a construction task are more likely to use both hands in a synchronized fashion and that this bimanual coordination predicted expertise. Similarly, studies have shown that learning gains can be
predicted by the amount of time students spend in certain postures (Schneider & Blikstein, 2015) and that the most productive posture involves both hands being synchronously engaged in the activity. This work was extended to look at dyad interactions, with the “driver” consistently using both hands more frequently while the “passenger” asynchronously moved their hands. Other MMLA work has shown that body posture during computer-supported activities can be predictive of learning (Grafsgaard et al., 2014) and theory suggests that increased body synchronization is associated with higher quality collaboration (Chartrand & Bargh, 1999). In summary, there is some emerging evidence that physical synchronization can be indicative of productive social interactions.

Visual synchronization

Eye-trackers have been used to study joint attention in collaborative learning situations. Richardson et al. (2007) showed that building upon a mutual source of understanding —“mutual grounding”— (i.e., hearing the same background information before the task) positively influenced the visual attention coordination in spontaneous discussions. More related to this particular study, Jermann et al. (2001) used synchronized eye-trackers to assess the degree of collaboration as programmers worked together on a segment of code. By a comparison of a ‘good’ and a ‘bad’ dyad, the study suggested that high joint visual recurrence is strongly related with collaboration. Nüssli (2009) showed that models of group behavior can be built with a combination of eye-tracking and other data: the combination of gaze and raw speech data (voice pitch and speed) afforded predictions of participants’ success with an accuracy rate of up to 91%. Lastly, Brennan et al. (2008) conducted a spatial search task and studied the effect of shared gaze and speech during the experiments; they concluded that the shared gaze condition surpassed solitary search by twofold in terms of speed and efficiency and was the most optimal of all the conditions. Consolidating the results from the above studies, we can see that joint attention and in turn synchronization between individuals are crucial for high-quality collaborations. The results suggest eye-tracking as a salient method for understanding factors that contribute to effective collaborations.

Physiological synchronization

Recently, researchers have started to study collaborative groups using electrodermal sensors. Electrodermal activity (EDA; also referred to as galvanic skin response, GSR) measures the amount of sweat produced by the sympathetic nervous system and is an indication of physiological arousal. By using synchronized EDA sensors, one can measure whether group members are aroused at the same time, or exhibit some levels of desynchronization. Early work by Pijeira-Díaz, Drachsler, Järvelä and Kirschner (2016) looked at different measures of physiological coupling indices (PCIs), and found that Directional Agreement (DA) predicted learning gains while Instantaneous Derivative Matching (IDM) was related to the quality of the produced artifact. In summary, there is some preliminary evidence that physiological synchronization can capture a facet of productive collaborations.

Summary

In sum, educational researchers are starting to use various kinds of sensors to capture facets of a productive collaboration. There is data suggesting that physical, visual and physiological synchrony can be used a proxy to collaboration quality. In the next section, we describe our study where we measured those states using a Kinect sensor, two mobile eye-trackers and two wristbands capturing participants’ electrodermal activity.

General description of the study

42 pairs of participants (N=84) programmed a robot to navigate a series of increasingly complex mazes (see for more details on the study, see Starr, Reilly & Schneider, 2018). Participants were shown two tutorial videos to acquaint them with the basics of block-based programming and how to use the values from sensors on the robot in their code. Groups were told to come up with a general solution that could solve any simple maze and then had 30 minutes to complete as many of the mazes as possible. Two different interventions were implemented to support collaboration in a two-by-two between-subjects design resulting in four different conditions (presence / absence of): 1) a visualization of the amount of individual verbal contributions as a proportion of total verbalization (Fig. 1, top left corner of the right picture; referred to as VISUALIZATION henceforth); 2) a short verbal explanation of the benefits of collaboration for learning (referred to as EXPLANATION henceforth). Outcomes of interest included the quality of the code groups produced (evaluated on a zero to four scale to determine how well the code in abstract could perform the maze solving task), the number of mazes solved, gains on a learning test administered before and after the sessions, and the quality of their collaboration. Multimodal data was collected via two mobile eye-trackers, two bracelets tracking electrodermal activity, a motion sensor, and video recording (Fig. 1).
Figure 1. A frame from the video used for the qualitative analyses of this paper. The two top images show the perspective of the participants (captured by the mobile eye-trackers) and the location of their gazes (red circle). The bottom left view shows the main video feed of the session, and the bottom right view displays a screen capture of the laptop. In this frame, group 8 was programming the robot to navigate an S-shaped maze.

### Data analysis: Contrasting group 7 and 8

The goal of this paper is to analyze two dyads in more depth and design measures that will allow us to contrast good versus poor collaborative styles across the entire sample. We chose to focus on groups 7 and 8 because of the stark differences in their behaviors. Group 8 (EXPLANATION, VISUALIZATION) was among the top groups in our sample: participants had a productive collaboration where group members built on each other’s ideas and exhibited positive learning gains. The participants in group 7 (EXPLANATION, NO VISUALIZATION), on the other hand, exhibited lower scores on all our metrics, resulting in periods of silence and miscommunications as well as negative learning gains.

#### Traditional quantitative measures

Group 7 was comprised of a 50-year old male (7L) and a 26-year-old female (7R). Both self-reported “Some College” for level of educational attainment and 7R indicated she was currently a student. 7L scored 8.3 percentage points worse on the post-test for computational thinking skills, indicating some confusion about the concepts required for the task. 7R gained 20.9 percentage points between pre and post, suggesting a much better grasp of the material after completing the activity. Participants in group 7 were able to direct the robot successfully through one maze but their final code failed to nest conditional statements and did not use the pre-written functions correctly (Fig. 2, left side). In a written reflection section on the post test, 7L had the following to say about his time with the activity: “…I have no talent for programming. I did not have any breakthrough moments. I would not do this type of study again. My ideas did not change over time and I felt that I did not learn much about computers.”

Group 8 was comprised of a 25-year old female (8L) and a 35-year-old female (8R). Both reported completing college and identified as no longer being students. 8L scored 16.7 percentage points higher on the post test of computational knowledge, indicating a modest improvement. 8R scored no better on the posttest compared to the pre, although she did make different errors. This suggests a level of confusion related to certain topics in computational thinking. They were also only able to complete the simplest maze but their code made use of nested conditional statements and correctly employed the prewritten functions they were given (Fig. 2 - right side). In a written reflection section on the post test, 8L had the following to say about his time with the activity: “We tried playing around with the different sensors. We started trying sensors 1 and 2, but then realized that using sensor 4 was necessary to complete the task. That was our "a-ha" moment. We tried changing the order of the if/else/do functions to get different results that helped advance our knowledge of the task.”
These two groups were selected due to their similarities of having one more knowledgeable participant, both completing the same number of mazes, and both having complete multimodal data to analyze. Exploring the difference in amount of collaboration observed and how that relates to the quality of the written code is a main goal of this work as well as identifying multimodal markers to signal quality of collaboration.

**Qualitative data**

Before delving into the overlay between qualitative data, eye-tracking data, and physiological data, we will overview the key themes seen between Groups 7 and 8’s qualitative data. To obtain this data, subjects’ experiments were recorded on video and an iterative process was used to note the various data points for qualitative data.

Group 7 engaged much less with each other than Group 8. There was little rapport built between the Group 7 members, whereas Group 8 laughed as early as the initial calibration color-reading exercise and continued throughout the experiment (17 times for Group 8 versus once for Group 7), especially after trials and during coding discussions. The length of dialogue in Group 7 was noticeably shorter than that of Group 8, and the content exchanged within Group 7 was not as detailed as Group 8. Even when Group 7 discussed more detailed code, it was a leader-follower response with the subject on the right saying most of the language and the subject on the left saying “mhmm” or “ok” and followed by heavy sighing. This type of dialogue impacted Group 7 negatively as the team was not able to understand the task at hand and dialogue came to a pause frequently through the experiment.

By contrast, Group 8 was almost the exact opposite of Group 7. Building on each other’s rapport minutes into the experiment, the two subjects were able to speak specifically about each section of the code. Each subject switched between making suggestions and verifying the assertions. The pair spoke at equal lengths throughout the experience, and often asked each other questions directly related to the task at hand. Unlike Group 7, which consistently displayed confusion via repeating statements like “I don’t know,” Group 8 displayed encouraging enthusiasm with short exclamations to relieve stress even as it recognized rising task difficulty throughout the experiment. For instance, about halfway through task 3, right subject suggested adding code to tell the robot to reverse direction, which is responded by an enthused left subject, “right!” In another instance, the robot did not make an intended turn, but turned in an opposite direction. The dyad exclaimed in surprise, but the right subject said, “well we got it to turn left, so that’s promising!” Lastly, Group 8 had many instances of “mhmm,” “ok,” and “does that make sense” language that were spoken by the dyad with friendly tone, versus the resigned tone of Group 7 for the same words. It is worthy to note that Group 7 did not have an equal split of reciprocal filler words, as the left subject were the one who said most of such language.

Group 8 also visibly and often celebrated for their successes, which did not happen for Group 7. Forms of celebration was most commonly displayed via loud exclamations like “whoo!”, high-fives, laughing and clapping. These observable signs provide grounds to believe that reinforcing signals of goodwill such as frequent check-ins and friendly body language and enthusiastic tones build rapport that help the team sustain collaboration as the task difficulty increases. Another key difference that set Group 8 apart is the frequency of the iterations. Building upon the alternating role of suggesting and verifying, the dyad was willing to make adjustments to the code, test it, and come back to the coding platform to improve the code if the trial failed. Combined with frequent acknowledgement of each other—including looking at each other—the dyad was able to
repeat the iterations numerous times, become increasingly familiar with the interface and be more hands-on than Group 7.

**Eye-tracking data**

Our first pass at analyzing visual synchronization involved applying the taxonomy developed by Kaplan and Hafner (2006), where they define joint visual attention as: simultaneous looking triggered by a salient event, simultaneous looking triggered by a “pop-out” effect, coincidental simultaneous looking, gaze following, or coordinated gaze on same object (Table 1). We stayed largely consistent with the five-category approach from Kaplan and Hafner (2006)’s hierarchy for generating our eye-tracking data. We decided to use this hierarchy because it was pertinent for gauging joint attention between the two agents—the partners in our study—and in turn the synchronization between the dyad. We used the ELAN software to track the eye-tracking data through annotating the occurrence and duration of the various categories of eye-tracking data as relevant data points appear in the experiment videos (Fig. 3). In the future, we will use automated ways of capturing joint visual attention using the fiducial markers tapped at various locations in the room.

**Table 1: Categories used in the eye-tracking data**

<table>
<thead>
<tr>
<th>Category</th>
<th>Hierarchy Shorthand</th>
<th>Example</th>
<th>Group 7</th>
<th>Group 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simultaneous looking triggered by a salient event</td>
<td>1 - Simultaneous looking_Salient</td>
<td>Agent 1 points finger at the screen, Agent 2 looks at the screen</td>
<td>19</td>
<td>52</td>
</tr>
<tr>
<td>Simultaneous looking triggered by a “pop-out” effect</td>
<td>2 - Simultaneous looking_PopOut</td>
<td>Agent 1 looks at a different color of code block because its color stands out</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Coincidental simultaneous looking</td>
<td>3 - Simultaneous looking_Coincidental</td>
<td>Agents 1 and 2 both looks for the robot, sees the robot at the same time but has no interaction with one another</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Gaze following</td>
<td>4 - Gaze Following</td>
<td>Agent 1’s gaze follows that of Agent 2</td>
<td>44</td>
<td>59</td>
</tr>
<tr>
<td>Coordinated gaze on same object</td>
<td>5 - Coordinated Gaze_SameObj</td>
<td>Both agents look at the same object knowing the other agent is looking at it as well</td>
<td>62</td>
<td>142</td>
</tr>
</tbody>
</table>

**Figure 3.** Screenshot of eye-tracking data for group 8 as seen in ELAN annotations, on the bottom half of the screen by category (i.e., second column of Table 1). Lengths of the annotated simultaneous gazes can be seen.

Overall, there were more coordinated gaze on the same object than other categories, followed by gaze following and then salient event. This is consistent with expectation for the study, as the dyad conducted most of its interactions via sedentary coding. The salient events were mainly caused by one subject pointing at the screen or the guide sheet provided. Gaze following usually occurred after salient events and before coordinated gazes. When gaze following overlaps with another category in sequence, the total combined gaze duration is shorter for the salient event than for the coordinated gaze.

For both dyads, most coordinated gazes longer than 10 seconds happened during task 3 (69% for Group 7 and 75% for Group 8), indicating the relative scale of attention required to complete a more difficult task. The salient event gazes did not last long for either group due to the nature of the gaze. What was different
was that Group 8 had a mostly even distribution of salient event gazes through the different task periods, while Group 7’s salient events were fewer and more clumped than of Group 8 (19 for Group 7 versus 32 for Group 8). Given that most of the salient events occurred because of finger pointing to attract attention of the subject’s partner, this suggests that Group 8 may have had more interactions with one another.

Additionally, eye-tracking results indicate further differentiation of Group 8 from Group 7. Group 8 had the longest joint eye gazing, and more often overall than Group 7 (142 coordinated gazes for Group 8 versus Group 7’s 62). Coordinated gaze on the same object composed of 53% of Group 8’s total eye-tracking data, and of that category 55% of the gaze durations surpassed 10 seconds long, compared to Group 7’s 11%. Finally, Group 8 had more overall eye-tracking appearances, with 266 tracked points to Group 7’s 133.

Comparing the relative frequency of appearance of each category across groups, Group 8 dominated across each category except gaze following. Group 7 had a great proportion of its eye-tracking events as gaze following when compared against Group 8. This suggests that simple gaze following may not contribute to the effectiveness of team collaboration--more active feedback (as measured by various types of eye movements and other data) had to be exchanged between the partners to create more meaningful interaction.

The above data suggests that Group 7 is not as synchronized as Group 8, with fewer coordinated gazes, fewer overall gazes, and shorter gazes than Group 8. In a task like pair programming, coordinated gazes are increasingly important as tasks become more complex. The drastic difference between Group 7 and Group 8 lead us to seek further validation in qualitative data to see if further patterns can be seen.

**Physiological data**

Tying the themes described above to EDA spikes, Group 7’s physiological arousals did not sync across the dyad and generally the state of physical arousal decreased throughout the experiment for one subject while for the other subject the EDA levels remained about the same (Fig. 4, left side). The spike in the EDA of the subject towards the right-hand side of the screen observed when the subjects sat and watched an instructional video. 40 seconds after the start of the activity, the left-hand subject had an EDA spike as he looked at the moderator when she was providing the pair directions. Neither of these events were related with the tasks at hand. The only relevant arousal happened 10 minutes after the beginning of the activity, when the right subject entered into an agreement period in which she was narrating the logic of the code to the left subject. The pair also looked at each other for the first time. Despite this, it’s clear that the left subject did not match the EDA state as right.

![Figure 4. EDA graph, normalized, for group 7 (left side) and group 8 (right side). Indices of synchronization can be computed using measures described by Pijeira-Díaz, Drachsler, Järvelä, & Kirschner (2016).](image)

Group 8’s physiological data showed that the dyad had a higher level of synchrony (Fig. 4, right side). In particular, about five minutes after the 5th tagging procedure, the dyad ran their robot through the first maze (task 2), and upon the robot’s success the right subject celebrated with a “whoa” and raised arms. This was picked up by the EDA as the intersection of the EDA measures of left and right subjects. The pair continued onwards through the remainder of the experiment in active EDA syncing, showing at least three visible intersection points starting from 20 minutes after the 5th tagging procedure, per Figure 4. The overlaps increased in frequency towards the end of the session, further supporting earlier eye-tracking and qualitative data showing the quality collaboration within the dyad. During this period, there were numerous iterations of coding and running the robot, often ending in laughter. The spikes in EDAs are mainly explained by these as well as the frequent standing by the dyad for their repeated robot retrials.

**Preliminary quantitative analyses**

In this section, we describe a first attempt at capturing physical and physiological synchrony between group 7 and group 8. Our strategy was to synchronize the data between group members and produce a scatter plot,
where values on the x-axis are shown for the first participant and values on the y-axis are shown for the second participant. We can then roughly subdivide a graph into four quadrants: the bottom left represents when both participants were exhibiting low levels of physiological activation or movement; the top right shows when they were both aroused or moving. The last two quadrants (top left or bottom right) indicate some levels of desynchronization: one group member has high values while the other participant has low values. Figure 5 shows that group 7 exhibits a pattern that is L-shaped as well as negative correlations (Fig. 5, left side), while group 8 tends to have points that are more evenly distributed - which is captured by positive correlations. We are planning to apply those measures to our entire sample to confirm those results.

<table>
<thead>
<tr>
<th>Group 7</th>
<th>Group 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EDA</strong></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="EDA Graph" /></td>
<td><img src="image" alt="EDA Graph" /></td>
</tr>
<tr>
<td>r = -0.047, p &lt; 0.05</td>
<td>r = 0.077, p &lt; 0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 7</th>
<th>Group 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kinect</strong></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Kinect Graph" /></td>
<td><img src="image" alt="Kinect Graph" /></td>
</tr>
<tr>
<td>r = -0.046, p &lt; 0.05</td>
<td>r = 0.179, p &lt; 0.05</td>
</tr>
</tbody>
</table>

**Figure 5.** Group 7 is on the left side, group 8 is on the right side. First row shows synchronized EDA data (the first participant is on the x-axis; the second participant is on the y-axis). Second row shows the amount of movement generated by each participant on each axis.

**Discussion**

In this paper, we contrasted two groups sampled from a larger study (N=84) using qualitative and quantitative methods. Qualitative analyses suggested that group 7 had more issues working together and accomplishing the task, while group 8 was more successfully and enjoyed the task more. The more collaborative dyad had much more detailed language and frequent, specific interactions, but also developed rapport through body language, mutual gazing, and frequent acknowledgement of each other. Combined with the use of keywords and tones that signaled positive intention, the more collaborative dyad was able to weather the stresses of completing difficult task and maintain task engagement as one unit. Physiological and eye-tracking data further validated our observational data, providing the multimodal view of team collaboration. Eye-tracking data showed that frequent simultaneous gazes and longer gaze length were characteristic of the more collaborative group, which eventually accomplished the tasks given. This provides support that eye-tracking data can provide a good gauge for high-quality group collaboration (as previously shown by Jermann, Mullins, Nuesli, & Dillenbourg, 2001). Gaze-following is also important in indicating reciprocity of the dyad in collaborative spaces. However, simple
gaze following is insufficient in contributing to collaboration. Additionally, electrodermal data confirmed those observations by showing elevated activation from the more successful group and helped us identify events of interest (e.g., by analyzing “spikes” in the data). Finally, those qualitative analyses helped us design measures of synchrony in small collaborative groups: we found that bodily and physiological synchronization should be further studied by extending the results found in Fig. 5 to the entire sample and confirming whether it can be used as a proxy for identifying successful groups.

Conclusion

In this paper, we have presented a new way of studying collaborative learning groups by using a combination of qualitative observations and data from high-frequency sensors. Our preliminary analyses suggest that Multi-Modal Learning Analytics (Blikstein & Worsley, 2016) can help us shed new lights in collaborative learning processes, especially in open-ended learning environments such as makerspaces. In the future, we are planning to further explore differences between groups in our study and develop multimodal proxies of collaborative interactions using high-frequency sensors. Those proxies could then be used by teachers and practitioners in informal learning environments to support the development of 21st century skills, especially in terms of students’ ability to work effectively in small groups.

References


Multimodal Texts and Tasks in Elementary Project-based Science

Miranda S. Fitzgerald, University of Michigan, mfitzge@umich.edu
Annemarie Sullivan Palincsar, University of Michigan, annemari@umich.edu

Abstract: This paper describes a study conducted in the context of a larger design-based research project investigating the integration of science, English language arts, and mathematics in elementary grades project-based learning. This study focuses on one third-grade unit in which students read and created multiple multimodal texts, in print and digital forms. Focusing on the design and enactment of two focal texts and associated tasks, we ask: (1) How did the design and enactment of the texts and tasks support third-graders’ science and literacy learning? (2) How might modifications to the design of the texts and tasks better support student learning? Findings indicate ways in which the design of the texts and tasks, and the teacher’s enactment, synergistically supported students’ science and literacy learning, but also point to missed opportunities in the design of the curriculum and its enactment, for further supporting the learning of all students.

Keywords: Multimodal literacy, science learning, designed learning environments, project-based learning

Introduction and purpose

Influenced by the growing range of digital technologies for communication, literacy scholars (e.g., Jewitt, 2008) have called for instruction that builds toward multimodal literacy; that is, the ability to learn from text in which words are used in combination with audio, visual, and spatial modes (Mills, 2010). Dalton (2012) argued that the Common Core State Standards (CCSS, 2010) “assume that being literate means being digitally literate” (p. 333) and that to be prepared for the 21st century demands of school, life, and work, students must be able to analyze and create both print and non-print texts using traditional and new forms of media (p. 333). However, some studies have found that, even when multimedia texts were recommended, elementary-grades teachers did not use these resources (e.g., Brenner, Hiebert, & Tompkins, 2009). Hence, studies that investigate the processes and outcomes of incorporating multimodal and digital texts into elementary-grades instruction are needed.

The present study was conducted in the context of a multi-year design-based research project that integrates science, English language arts (ELA), and mathematics in elementary grades project-based learning (PBL), called Multiple Literacies in Project-based Learning (MLs). PBL hypothetically provides opportunities for students to interpret and produce multimodal text as they explore real-world problems. Project-based science instruction, in particular, may provide unique opportunities to explore how students and their teachers use, create, and learn with multimodal text because, as Lemke (2004) argued, scientific literacy is inherently multimodal and scientific disciplines are “leading the way” in the use of video, graphical displays, and simulations to pursue research questions.

While project-based science may provide opportunities for students to interpret and produce multiple modes, we know very little about how young students and their teachers take up these opportunities in diverse classroom settings. Prain and Waldrip (2006) found that, while teachers incorporated multiple modes into science instruction, they did not systematically support students to translate across modes, and students required varying levels of support to interpret modes. This raises questions about how young students and their teachers can be supported to interpret and produce multiple modes of representation as they engage with disciplinary texts and with one another. The current study is consistent with the ICLS 2018 focus on unpacking the complexity of the learning and teaching process, exploring learning in real-world settings, and understanding how learning may be facilitated with and without technology. Reflecting this call, design-based studies are needed to understand how multimodal texts and tasks can be designed and used to support students’ knowledge building and literacy development.

This study was conducted in the context of one third-grade project-based unit, framed by the driving question: Why do we see so many squirrels but we cannot find any stegosaurus? The unit was enacted in a diverse third-grade classroom, in which the class studied organisms’ traits and interdependent relationships in ecosystems. The unit featured multimodal texts and tasks designed to advance students’ literacy and science learning. To investigate the design and enactment of the texts and tasks, we ask: (1) How did the design and enactment of multimodal texts and tasks support third-graders’ science and literacy learning in the context of PBL? (2) How might modifications to the design of the texts and tasks better support third-graders’ science and literacy learning?
Theoretical perspectives

The RAND Reading Study Group (2002) proposed that reading comprehension can be explained by considering the reader, the text, the activity in which the reader and text are involved, and the sociocultural context in which the activity takes place. This study focuses on students’ interpretation and knowledge building in the context of reading, interpreting, and creating multimodal texts. This work is also informed by sociocultural theories of learning, which reject the view that knowledge is located within the individual and, instead, embrace the view that learning and understanding are inherently social, occurring through interaction, negotiation, and collaboration (Wertsch, 1991). The present study’s focus is on the classroom community as the students and their teacher interact and collaborate around reading, interpreting, and creating multimodal text. From sociocultural perspectives, cultural activities (e.g., scientific modeling) and tools (e.g., computers, language) are integral to knowledge building.

Methods and data sources

This study was conducted in one third-grade classroom with 32 students in a K-5 elementary school in the Midwest United States. Students were diverse with respect to race/ethnicity and academic achievement. On state achievement measures, only 20% of students demonstrated proficiency in ELA. The teacher was an experienced third-grade teacher and a second-year participant in the MLs research project.

Data sources included field notes, transcribed video of instruction, and artifacts. Informed by Brenner et al. (2009), observations and qualitative field notes focused on instructional events in which students were engaged in reading and using information from text to: discuss ideas, inform the development of artifacts, and synthesize ideas across texts and experiences. Focal lessons were videoed and transcribed. Artifacts included curriculum materials and student and whole-class work. Artifacts provided evidence of how students incorporated ideas from text in student- or class-generated products and provided insight into students’ knowledge building and literacy development.

We use design-based research (DBR) (Brown, 1992) and case study (Stake, 1995) methods focused on the design, evaluation, and improvement of an intervention as it interacts with the contextual variables integral to enactment (Fishman et al., 2004). In this study, we analyzed the design and enactment of a set of texts and tasks in one MLs unit of instruction. Through analyzing multiple data sources and constructing in-depth cases, our goal was to uncover the features of the design and enactment of texts and tasks to identify how these supported or failed to support knowledge building and literacy development. Identification of missed opportunities and modifications provides evidence to support design revisions that might better support student learning in this context.

Analyses began with reading through transcript data line-by-line to identify relevant episodes of instruction and applying notes and codes responsive to our research questions. The next phase of analysis involved direct interpretation (Stake, 1995) of episodes within transcripts, and use of connecting strategies (Maxwell, 2013) to construct cases. This holistic examination of transcript episodes, designed curriculum, and related artifact data allowed us to uncover the ways in which the focal lessons unfolded – specific to the design and enactment of texts and tasks, and the ways in which students took up these opportunities – in order to develop assertions that were responsive to our research questions. Analyses supported our construction of narrative summaries to develop our case study report, in which we sought to maintain both the context and story of the ways in which instruction unfolded in the classroom during focal lessons (Maxwell, 2013).

Instructional context: The unit of instruction

To launch the unit, the teacher introduced the driving question and students viewed a video clip of Jurassic era organisms to identify similarities and differences between the organisms in the video and those that students see around their homes and school. Students then conducted observations of squirrels and other organisms in their habitats. Based on observations, students created initial models to explain how squirrels survive. In science, models are used to represent a system under investigation, develop explanations, make predictions, and communicate ideas (NGSS, 2013, Appendix F). Because this was students’ first experience with modeling, their models took the form of concrete drawings of squirrels in their habitat (NRC Framework, 2012).

In the next set of lessons, students investigated squirrels’ structures (e.g., teeth, tail, legs) by observing photographs, videos, and a squirrel’s skull. The teacher then read aloud a text designed to introduce, describe, and engage students in making close observations of structures that enable squirrels to climb headfirst down trees. The researcher-designed text was paired with a video clip, illustrating the ideas in the text. After reading, the students returned to their initial models to make revisions to reflect new learning. After revising, students explored structure-
function relationships of – and interactions among – organisms that share habitats with squirrels. Students selected and read one text from a set of researcher-designed texts, which provided written and visual information about a set of organisms found in squirrel’s habitats (e.g., ant, coyote, tree). In small groups, students read and answered questions about the organism, drew and labeled organism structures, and identified interactions with other organisms. Students then used this information to co-construct a class model to explain interactions among the organisms in an ecosystem.

Findings

In this section, we describe findings from our analyses of the design and enactment of two sets of texts and tasks.

For Squirrels, It’s Headfirst and Down and paired video: Design and enactment

*For Squirrels, it’s Headfirst and Down* was designed to: (a) illustrate core ideas related to adaptations that enable organisms to survive in particular habitats, (b) provide information that built upon students’ first-hand observations, (c) include multimodal features (written text and images), and (d) motivate and provide new information to support engagement in scientific modeling based on new learning. The paired video was selected to illustrate the ideas in the text and to provide an opportunity for students to observe a squirrel using the structures depicted in the text and photographs. We found that the design of the texts and tasks, and the teacher’s enactment, synergistically supported third-graders’ science and literacy learning. We also identified missed opportunities, within the design of the curriculum and enactment, for further supporting the science and literacy learning of all students.

Before reading: Setting the purpose and preparing to read

The teacher leveraged the text and task to support students’ science and literacy learning by supporting students to: (a) connect to prior knowledge and experiences, (b) identify the purpose of the text and task, and (c) make predictions prior to reading. The teacher began the lesson by asking students to recall their recent experiences in the PBL unit, which included priming students’ thinking about squirrels’ structures investigated during previous lessons using multiple modes of representation (photographs, videos, skeletons). Specific to literacy learning, supporting students to activate prior knowledge before reading creates opportunities for students to draw on related knowledge and experiences to interpret new information in text (Brown, Pressley, Van Meter, & Schuder, 1996). In this case, students’ prior knowledge was shared knowledge the class developed through analyzing multiple modes of representation.

One purpose for reading this text was to provide new evidence that students could add to their models, in order to explain squirrel survival. After prompting students to think about their earlier observations and investigations, the teacher explained, “We’re going to read this together, but as we’re reading, I’m going to ask you some questions. And then we’re going to…look at our models.” This excerpt illustrates how the teacher connected the purpose of the reading to the modeling task. In research investigating teachers’ differential enactments of an inquiry curriculum, Puntambeker, Stylianou, and Goldstein’s (2007) findings emphasized the teacher’s role in helping students understand connections among instructional activities and concepts in order to drive student learning.

During reading: Supporting students to read and interpret information

During the interactive read aloud, the teacher supported students’ reading and interpretation by: (a) engaging students in visualizing and acting out the ideas in the text, and (b) checking for understanding. After reading a section that described how squirrels use their sharp claws to grip a tree (“With its sharp claws, a squirrel can grip the bark of a tree. The strong grip of the front claws allows the squirrel to hold on while it moves its back feet. Then, the back feet can hold on while the front feet move.”), the teacher asked students to “picture” the description in their mind. The teacher reread this portion of the text and made several gestures, which multiple students imitated, to demonstrate the squirrel’s movement. This echoes Glenberg’s (2010) findings that one way to enhance reading comprehension is by supporting students to make connections between the text and its embodied meaning.

The teacher also paused frequently to check for understanding as the class read the text aloud together. Her questions focused on the meaning of the words in the text as well as supporting students to analyze and interpret information in photographs. In one example, the teacher paused and, as prompted in the text (“Look closely at the back feet of the squirrels in these photographs.”), asked students to closely observe two photographs.

Teacher: [Reading aloud] “To find out about the last feature that helps squirrels to climb down the tree headfirst, look closely at the back feet of the squirrels in these
photographs.” … you should be on those two pictures…they’re telling you to look closely at what?

Kaylee: The pictures!
Student: The squirrels.
Student: The feet.
Student: The back feet.
Teacher: The back feet. They’re specifically telling you to look closely at the back feet… What did you notice about those back feet, Jessica?
Jessica: I noticed that the tree one is actually like hanging upside down on the tree and the other claws are just leaning on it.
Teacher: Okay, so it’s hanging upside down… What else do you notice about him, Aiden?
Aiden: That I notice that it’s not exactly hanging down. It’s actually using its back feet to push off and using its front feet also to push off.

In this excerpt, the teacher asked students to study the photographs and invited them to share their observations. The teacher then guided students to clarify what, specifically, they should examine in the photographs (i.e., squirrels’ back feet). After clarifying, the teacher invited students to share their observations before continuing to read the text.

**After reading: Using information from the text to inform revisions to scientific models**

The teacher used the text and task to scaffold students’ science and literacy learning by supporting students to identify and use text ideas to revise models in whole-class, small-group, and one-on-one contexts. The teacher also leveraged the text and task to scaffold students’ understanding of and engagement in scientific modeling.

After reading, the teacher made a number of instructional moves to scaffold students’ use of the information in the text to inform revisions to their models. For instance, immediately after reading, the teacher prompted students to apply the reading to the modeling task and asked students to summarize the structures that squirrels use to climb headfirst down trees. The teacher then played a video clip of a squirrel climbing and leveraged this as another opportunity to illustrate the structures students read about using a different mode of representation (e.g., Teacher: “Look how its…back leg is completely turned around.” Brody: “I can see it!”). Once students began revising their models, the teacher scaffolded their use of information from the text, photographs, and videos by conferring with individual students, inviting them to share their revisions with the class, and guiding them to integrate visual and written modes of communication with their models.

Teacher: Oh, Ellie’s adding something… Ellie, what are you adding?
Ellie: I’m adding…another squirrel in the position that it can be.
Teacher: Okay, you’re adding another squirrel, and you’re putting it in what kind of position? Can you describe that for me?
Ellie: In the position going down where its legs are facing backwards and its front paws are facing forward.
Teacher: …what else could you do in your models once you’ve drawn that squirrel…?
Sam: Label.
Teacher: Label it. Label the back legs. What specifically would you label, Sam?
Sam: Claws.
Teacher: Okay, it has claws. What else does it have? Aiden?
Aiden: It has an anklebone.
Teacher: It has an anklebone, right? Those are the structures that allow it to go headfirst and down.
This excerpt illustrates how the teacher used an individual student’s model to launch continued review and discussion of the structures introduced in the reading. Also, one purpose of scientific models is to communicate ideas to others (NGSS, 2013, Appendix F). This example illustrates how the teacher used Ellie’s model to introduce the idea of combining labels and drawings to clearly communicate science ideas. Figure 1 (a) shows a portion of Ellie’s model. After the above exchange, Ellie added labels for the squirrel’s anklebone and claws.

![Figure 1](image)

**Figure 1.** Ellie’s model (a) includes drawings and labels. Jenna’s (b) and Malik’s (c) models provide more examples.

On the second day of enactment, the teacher continued to scaffold students’ use of the information in the text to inform revisions to their models by (a) sharing and discussing additional students’ models, (b) frequently directing students’ attention back to information in the text, and (c) clarifying the purpose of the modeling task.

**Constraints of the text and task revealed through enactment**
Specific to our second research question, one limitation was the written curriculum’s lack of guidance for engaging students in text-based discussion. While the teacher frequently paused to check students’ understanding of images and words, many questions did not elicit high-level thinking, which is important for supporting comprehension in the context of discussion (Soter et al., 2008). Further, both the written curriculum and the teacher’s enactment missed opportunities for students to integrate ideas in the text with related activities, such as investigations and observations, done prior to reading. This echoes Prain and Waldrip’s (2006) finding that teachers did not systematically support students to translate across modes. These findings suggest that teachers may benefit from additional support to enact text-based discussions, particularly when reading tasks call for translating across modes and making connections to related unit activities. In revisions, we have begun to design more interactive reading guides (Arias, Palincsar, & Davis, 2015) to support students’ discussion and sense-making with multimodal texts in our PBL units.

**Structure-function cards and paired map: Design and enactment**
The organism structure-function cards were designed to: (a) illustrate core ideas related to adaptations that enable organisms to survive in a particular environment; (b) provide information that connected to and built upon students’ first-hand observations within and beyond the classroom; (c) include multimodal features (written text and images), and (d) motivate and provide information to support students’ engagement in the practice of scientific modeling. In the focal classroom, students used digital versions of the structure-function cards, accessed via Chromebooks. Also, a paired digital map (Google Map of the area around the students’ school) was selected to engage students in using their class interactions model to make predictions about organisms that live and interact in a local context. We found that the design of the texts and tasks, and the teacher’s enactment of the organism structure-function cards synergistically supported third-graders’ science and literacy learning. Again, we found missed opportunities, both within the design of the curriculum resources and the enactment of the lessons, for further supporting science and literacy learning.

**Before reading: Setting the purpose and preparing to read**
Analysis of the written curriculum and transcripts of classroom enactment revealed that the teacher leveraged the texts and task to support students’ science and literacy learning prior to reading by: (a) supporting students to activate prior knowledge, (b) clarifying key vocabulary, and (c) setting a clear purpose for reading.

**During reading: Supporting students to read and interpret information**
The teacher leveraged the texts and task in support of students’ science and literacy learning by supporting students to read and interpret information in the texts. Students’ partner reading and interpretation of their chosen structure-
function text was scaffolded by the guiding questions included in the curriculum materials and the teacher’s use of the questions as she conferred with individual and small groups of students about their reading and learning with text.

During reading, students worked with a partner or group to respond to a set of guiding questions as they read text and analyzed images about their organism. The guiding questions were: (1) What is your organism? (2) List the structures you read about. (3) Choose one of the structures you listed. How does this structure help the organism meet a basic need? (4) Does your organism live in the same environment as squirrels? (5) What foods does your organism need to survive? (6) What predators eat your organism? These questions were designed to support students to identify information about how their organism interacts with other organisms in its habitat, in order to support the co-construction of the organism interactions model. In one-on-one, small group, and whole-class conversations, the teacher used the guiding questions to support students to read and interpret multiple modes (i.e., text, photographs).

Analysis of students’ responses to guiding questions provided additional evidence of how the questions scaffolded students’ reading and interpretation of the organism texts. While ten of sixteen focal students completed all six questions, all focal students identified their organism, listed structures, and described how one of those structures enables the organism to survive. The NRC Framework (2012) explains that, “The quality of a student-developed model will be highly dependent on prior knowledge and skill and also on the student’s understanding of the system being modeled” (p. 59). In other words, the quality of the class-developed interactions model depended on students’ prior knowledge and understanding of the environment as a system, including the organisms that are a part of that system. We argue that, in hand with students’ firsthand observations and investigations, the multimodal information in the structure-function cards and the scaffolding provided by the guiding questions supported students to build the prior knowledge necessary to co-construct their organism interactions model as a class.

**After reading: Using text information to inform construction of the interactions model**

After reading, the teacher used the lesson plans, texts, and task to scaffold students’: (a) sharing of learning from text to develop shared knowledge, (b) use of shared knowledge to co-construct the interactions model, and (c) application of the model to explain organisms’ interactions in a local context. The teacher provided opportunities for students to share what they learned about their selected organism with the rest of the class. This was important because, in order for students to co-construct a model to explain how organisms in the environment interact, the students needed to build shared knowledge about the organisms. After responding to the guiding questions, the teacher invited students to debrief their findings. To conclude this day of enactment, two groups had the opportunity to: (a) present their learning by sharing their responses to the guiding questions and also (b) show the class how they navigated and identified information in the digital text (e.g., tapping to select the text, scrolling to locate information).

On the next day of instruction, the teacher introduced the *interactions* modeling task, and students used what they had learned about their organism to create a small drawing, on which they labeled one or more of the organism’s structures and described how it interacts with squirrels. After completing this step, each group shared and taped their organism card on the chart paper where the class would co-construct the interactions model. After attaching the cards to the chart paper, students had additional opportunities to share their learning about the organisms’ structures and how the organism interacts with squirrels (e.g., the coyote is a predator of the squirrel). Transcripts of whole-class discussions revealed that the students were developing shared knowledge of the different organisms, structures that enable them to survive in their environment, and ways in which they interact with the squirrel and other organisms.

The teacher supported students to leverage their learning from the readings to co-construct the interactions model, and asked students to suggest ways they could represent and communicate interactions among organisms. Initially, students proposed that they could “look on the Chromebook” to show relationships,” “act it out,” or “write it.” The teacher pressed students to think of ways they could clearly communicate the relationships among organisms directly on the chart paper. The design of this modeling task and the way in which the cards were arranged on the chart paper called for students to create a more abstract representation than the previous squirrel environment models, in which students drew and labeled concrete pictures. As the teacher pressed for more ideas, one student proposed that they could draw and label lines between organisms to communicate the nature of their relationships (e.g., predator/prey, shelter, etc.), a proposal which was taken up by the class.

Zayn: Line with the squirrel to the red-backed salamander. The squirrel is a predator…

Teacher: The squirrel is a predator to the red-backed salamander…what’s the special structure that the salamander has that allows it to get away…?
Zayn: Its tail.
Zayn: Falls off so it can run away, and once it run away, far away, it grows a new one.

This excerpt illustrates the ways in which the teacher enlisted students’ contributions in order to co-construct the interactions model to explain relationships among organisms in a habitat. Understanding and being able to explain the organisms’ unique structures and their functions, supported students to co-construct representations of complex interactions among the organisms included in the model (see Figure 2, below).

Figure 2. Co-constructed interactions model prior to and after identifying and labeling organisms’ interactions.

After constructing the interactions model, the teacher extended the written curriculum by engaging students in analyzing a digital text – a satellite view Google Map of the area around their school – and in using their model to make predictions about where on the map students might find squirrels and why. Prior to asking students to use their model to apply the phenomenon under study, she first supported students to interpret the map more generally (“Tell me what this is a map of. How do you know? What do you see?”). While the digital satellite map was an unfamiliar mode to students, they were able to draw on prior knowledge of landmarks and buildings because the map represented an area with which all students were familiar. After completing an initial “reading” of the map, the teacher asked students to make connections between the model and the environment around their school (“Based on our interactive map [and] based on what we know about squirrels and their environment…where do you think we would be able to find the most squirrels?”), and supported students to use evidence from the model and their firsthand experiences to make predictions about organisms they might find in the area (e.g., frogs, coyotes, etc.).

Constraints of the text and task revealed through enactment
One limitation of the written curriculum, revealed in analyses, was the lack of science content knowledge supports for the teacher to facilitate enactment of the organism text-reading and modeling task. Analyses revealed two instances, in the context of clarifying vocabulary before reading, when enactment may have contributed to confusion among students about what does and does not qualify as an organism. One potential contribution of providing educative curriculum materials, such as interactive reading guides, is to enhance teachers’ content knowledge (Arias, Bismack, Davis, & Palincsar, 2016). Providing this type of information in future iterations of the curriculum may better support teachers to scaffold students’ understanding of science vocabulary and concepts prior to, during, and after reading.

Significance
This work is significant because, as Lemke (2004) urged, students need to be able to interpret, analyze, and produce multiple modes of representation in service of disciplinary knowledge building. Aligned with the 2018 conference theme, this work also helps us to understand the grainsize at which we need to study teachers’ practice in order to unpack the complexity of the learning and teaching process. In the present study, this complexity is revealed through analyzing teacher discourse, teacher-student interactions, and students’ use and development of multimodal texts. Additionally, characterizing students’ opportunities to learn requires the close analysis of teaching and learning in real-world settings (Litman et al., 2017). This kind of analysis can support curriculum designers to engineer
opportunities to learn that are different from those typically offered in K-12 classrooms, and to more fully support teachers in managing the complexity of maximizing all students’ opportunities to learn in diverse classroom settings.

References
RAND Reading Study Group. (2002). Reading for understanding: Toward an R&D program in reading comprehension. Santa Monica, CA: RAND.

Acknowledgements
The research and development reported in this paper is being supported by a generous grant awarded to Joseph Krajcik, Annemarie Palincsar, and Emily Miller from the George Lucas Educational Fund entitled, Multiple Literacies in Project-based Learning. The research is further supported by a generous grant awarded to Miranda Fitzgerald from the AERA Division C Student Research Grant. The authors gratefully acknowledge Emily Miller, who led the design of this unit of study. Any opinions, findings, and recommendations expressed in this paper are those of the authors.
“Ohhh, Now I Can Do It!”:
School-age Children’s Spontaneous Mathematical Sensemaking in Construction Play

Lara Jasien, Vanderbilt University, lara.jasien@vanderbilt.edu
Ilana Seidel Horn, Vanderbilt University, ilana.horn@vanderbilt.edu

Abstract: This analysis joins together two lines of work: mathematical problem solving and children’s construction play as a resource for mathematics learning. Our study is motivated by two observations. First, play has characteristics reminiscent of professional mathematicians’ practice. Second, the child-centeredness of play points to possibilities for equitable mathematics instruction. We conceptualize problem solving in construction play as the process children engage in when they experience trouble accomplishing a goal and work to repair it. To examine this phenomenon, we used head-mounted GoPros to collect 348 point-of-view videos of children as they played in a mathematical playground, where “messing around” with designed objects provided them with opportunities to encounter mathematical concepts. Through a close examination of one case, we illustrate the mathematical generativity of play with the designed objects and argue that children’s construction play can support rich opportunities for children’s mathematical sensemaking.

Keywords: mathematics learning, out-of-school learning, play-based learning; mathematical sensemaking; interaction analysis

The heart of children’s play involves asking what if questions about the world. Similarly, the heart of mathematical activity involves asking what if questions about patterns in number, shape, and space. Thus, we posit that children’s play in mathematically rich contexts provides a productive avenue for them to conjecture and explore in disciplinarily authentic ways. Yet little is known about how school-age children might make sense of ideas in mathematically-rich play settings.

In the U.S., children’s opportunities to encounter mathematics are largely limited to classrooms. This is a significant shortcoming, since many children’s school mathematics experiences emphasize ritualized activity and pre-determined solutions (Stigler & Hiebert, 2009), often resulting in them disliking the discipline (Boaler & Greeno, 2000). This type of instruction is at odds with authentic disciplinary engagement (Engle & Conant, 2002) and mathematical practices like exploration, defining, and conjecture-testing (Lakatos, 1978).

This is a missed opportunity, since research has shown that self-selected and self-directed out-of-school learning helps students develop positive disciplinary identities and builds foundations for formal knowledge, even in technical fields like science, technology, and engineering (Quinn & Bell, 2013). Out-of-school settings facilitate children’s focus, persistence, joy, and pride in work (Petrich, Wilkinson, & Bevan, 2013). While out-of-school studies in museums and makerspaces have investigated students’ scientific, technological, and engineering activities, informal mathematics remains underexamined, offering little guidance for designing spaces that might support mathematical exploration.

We study out-of-school mathematics learning by investigating mathematically generative play in a mathematical playground called Math On-A-Stick (MOAS). To identify mathematically generative play, we look for moments when what if questions of play intersect with what if questions about disciplinary concepts. Specifically, we examine children’s construction play — play with materials that require building or designing. In construction play, children explore objects’ properties and see what they can do with them. In the process, they experiment and make tacit conjectures around patterns in number, shape, and space, which we call mathematical generativity. In addition to opportunities to do mathematics, mathematically generative play opens up possibilities for students to experience disciplinary enjoyment (Sengupta-Irving & Enyedy, 2015) of mathematics, potentially supporting their long-term identification with and persistence in mathematics.

Our study has shown that construction play is rich with opportunities for informal mathematics learning (Seo & Ginsburg, 2004), yet play of any kind is increasingly being pushed out of U.S. preschool, elementary, and middle schools due to increased emphasis on standardized tests and standardized curricula (Bodrova & Leong, 2003; Jerret, 2015). This situation poses a dilemma for teachers who wish to incorporate mathematical play into their classrooms, with thin empirical basis for its value. Eventually, we hope our research will influence classroom instruction, but
our first step is to examine how meaningful mathematics is joyfully engaged in school-age children’s construction play.

**Prior work: Construction play and children’s mathematical learning**

Prior research on how mathematics is engaged through play highlights construction play as mathematically generative (Sarama & Clements, 2009; Seo & Ginsburg, 2004). In construction play, children design with objects and engage in higher-level thinking as they solve problems that emerge from the constraints of construction materials (Bergen, 2009; Forman, 2006). There is evidence that preschool children who engage in construction play might do better in math later on (Stannard, Wolfgang, Jones, & Phelps, 2001). Additionally, older children who do well on construction tasks also tend to perform well in mathematics (Casey, Pezaris, & Bassi, 2012). While most of these studies were conducted in laboratories without intentionally designed materials, pedagogical theorists such as Maria Montessori have made much work of these ideas. In fact, Montessori schools in the U.S. leverage exploration in construction play as a primary aspect of their preschool curriculum. These pedagogies specifically leverage design of materials to promote scripted mathematical exploration.

However, we break from Montessori both in the age of children studied, and in what we count as play. First, the mathematical sensemaking of young children likely looks very different than the mathematical sensemaking of older children (Clements & Battista, 1992). Second, the nature of scaffolding in Montessori—where teachers model how to engage with materials (Lillard, 2013)—runs counter to our notions of play. Our conceptualizations of play feature children’s agency in exploration, self-selection of goals, and self-direction in how to accomplish them (Wing, 1995). Just as mathematics education research highlights how agency benefits learning, we hypothesize that children’s agency in play will lead to rich exploration of construction play materials. We argue that examining children’s self-directed construction activity with mathematically structured objects is likely to provide rich examples of children’s emerging mathematical sensemaking.

Surprisingly, there are few studies of upper-elementary and middle school aged children’s construction play with materials explicitly designed to direct children’s attention to mathematical concepts—perhaps because foundational scholars like Piaget (1946/1962) and Vygotsky (1978) focus on sociodramatic play as the leading activity of this age group. This has created a missed opportunity, as engagement with sophisticated materials can lead to productive mathematical exploration for older children (Papert, 1980). In fact, the mathematical generativity of construction play emerges from the material constraints that lead children to experience trouble as they work to achieve their goals (Bergen, 2009; Papert, 1980). Our research thus asks: How do school-age children engage in mathematically generative play in a designed out-of-school mathematical playground?

**Conceptual framework: Mathematically generative play as problem-solving through trouble and repair**

**Theoretical framework**

To understand how upper elementary and middle school children engage in mathematically generative play, we follow Vygotsky’s (1978) theory of mediated action. This theory holds that all knowing and doing is mediated by cultural tools, including language and physical objects. The objects at MOAS were intentionally designed to facilitate encounters with mathematical concepts (Wertsch, 1998). When the what if aspect of play intersects with the mathematical practice of conjecturing, we posit that children are working just beyond what they already know and understand. Questions like, what if I move this tile here? what if I rotate it this way? support explorations and encounters with mathematical concepts. When this happens, we see this as evidence of children operating in their zones of proximal development (ZPDs; Vygotsky, 1978), making these moments of potential learning. Of particular interest are moments where children (a) encounter trouble in meeting their self-generated construction play goals and then (b) make multiple attempts to repair that trouble, whether independently as the materials themselves act as scaffolds or with a more expert other who poses questions or models solutions. We consider these episodes of trouble-and-repair to be examples of problem-solving, as children’s work in their ZPD inherently involves problematic experiences for children relative their current understandings.

**Mathematical learning in play**

Foundational to this study is a view of math embedded in play (Ginsburg, 2006). Notably, Seo and Ginsburg (2004) looked at young children (<5 years) playing with mathematically rich objects and inductively developed
explanatory categories for their activity. In doing so, they identified mathematical ideas children could notice, even when children did not frame their activity as math.

To understand children’s sensemaking in play, we attend to children’s problem-solving during goal-based play. This is both a theoretical and a methodological choice. Theoretically, problem-solving requires an end goal, so identifiable goals help us analyze it. Methodologically, we count construction play to begin when children set goals (Hutt, 1979; Vandenberg, 1980), and these are often most visible when they are not met. We acknowledge that (a) children engage in exploratory play without construction goals, (b) children have goals we cannot identify, and (c) children may achieve their goals without experiencing trouble and may be engaging important mathematical ideas. Pragmatically, we analyze the cases where the problem solving is most visible and omit these cases where it is less so. As a result, we focus only on episodes in which children set visible goals, experience trouble, and make multiple attempts at repair, as these are low-inference cases for the phenomenon of interest.

Research design: Understanding problem solving and mathematical generativity in play

To explain our logic of inquiry, we describe our data collection procedures and identify our units of analysis. We then outline our data sampling and analysis methods for our research question, how do school-age children engage in mathematicaly generative play in an out-of-school mathematical playground?

Site selection

Research took place at a mathematical playground called Math On-A-Stick (MOAS) at the Minnesota State Fair. Data were collected over the full ten-days of the fair in 2016. MOAS was open to any fairgoer. It was a pleasant, shady space in a relatively quiet corner, containing nine tables, each of which had unique mathematically structured objects that children could use as they pleased along with volunteers who could facilitate play. These objects included pattern machines, various tiling pentagons, tessellating turtles, and 6x5 egg crates with colorful plastic eggs. Children’s voluntary participation, their freedom in how (and how long) to engage with the exhibits, and the mathematical richness of their design support a study of children’s sensemaking as they encounter mathematical ideas in unexpected ways.

Data collection

To understand the mathematical generativity of children’s play, data collection aimed to capture children’s perspectives of their activity in MOAS exhibits. We recruited 348 children to participate. They answered brief intake surveys and then went about the playground with Go-Pro video recorders mounted on their heads, aimed downwards and slightly forward to capture the children’s and adults’ talk, gestures, and object manipulation. This view captures the locus of the children’s attention when they are playing with objects: When the child looks up, the camera shows what they are looking at. Even slight glances at other children playing are captured. This video record supports inferences about children’s attention and interest. As participants exited MOAS, we conducted brief interviews about their experiences. The video data serve as the primary data for the study, with the survey and interview data used primarily to select participants based on different reported characteristics and experiences.

Data corpus

The 348 participants ranged in age from 4 to 16 years old. The average MOAS visit lasted 26 minutes ($sd = 0.007$), with median visit per exhibit at approximately four minutes. Out of the 348 participants, this study focuses on upper elementary and lower secondary aged children’s play ($7 – 12$ years old, $n = 277$, visit length: $m = 28$ minutes , $sd = 0.01$). This limits the variation in children’s development and addresses a gap in the literature, as this age group’s out-of-school mathematical learning is understudied.

Case selection

To illustrate how our methodology supports an analysis of mathematically generative play, we present the case of one child at the egg exhibit at MOAS. Olivia (a pseudonym) was an eight year old girl who played at the egg exhibit for over 12 minutes, much longer than the average stay time. We conjectured that sustained engagement at one exhibit might support increasingly complex design goals. We focus on Olivia’s last construction at the egg table, where she had a goal of making a heart design using pink plastic eggs in a 6x5 egg crate. This episode of goal-based activity lasted approximately seven minutes, meaning that this single design took her longer than
most children remained at this exhibit. Because Olivia’s mother was attentive while she engaged in construction play, their interaction was analytically useful, because it gave us access to Olivia’s thinking.

Unit of analysis: Episodes of goal-based activity
Because we seek to identify mathematically generative play, we distinguish between (a) play that produces mathematical ideas that the children might have noticed and (b) play that produces mathematical ideas that we can empirically argue that they noticed. In other words, while we can certainly infer mathematical ideas from children’s play activities, we seek to identify the mathematics that children attend to in play, even if it is not formally named.

As a preliminary data reduction strategy, we focused on episodes of goal-based activity, parsing our subset of videos into these episodes. To allow enough time for exploratory play and to seek examples of sustained engagement, we sampled video from children who stayed at an exhibit at least five consecutive minutes. Out of the 277 school-aged children in our sample visiting the many exhibits at MOAS, this data reduction yielded 395 episodes of exhibit level activity (~50 hours of video data). We reasoned that longer episodes were more likely involve multiple attempts at reaching the goal. To determine the start of goal-based activity, we attended to both explicit goals (e.g., the statement “I’m going to make a dog!”) and implicit goals (e.g., making a “blank slate” by ruining what was previously made). To determine the end of goal-based activity, we attended to when children: (a) discard (e.g., child destroys what made); (b) abandon (e.g., walks away without finishing); (c) preserve (e.g., having parents take photographs), or (d) share (e.g., soliciting adult praise). If, after preserving or sharing, the child continued to work on the same object to refine (but not dramatically alter) their work, we considered this part of the same goal-based activity, reapplying the criteria for the end of goal-based activity to determine when the activity truly ended for the child.

Second, we further reduced our data by identifying episodes of trouble-and-repair. Because we are interested in the mathematical learning supported by play, these moments made visible times when children had trouble reaching a goal, along with their strategies and resources for repairing the trouble. Trouble was located through children’s: (a) repeated attempts; (b) use of multiple strategies; and (c) expressions of frustration or confusion. While some exhibits afforded longer play than others, analyses show that two minutes of goal-based activity often captures multiple attempts at repairing a problem. We selected Olivia’s case by examining the duration of episodes of trouble in our data set, with hers being and outlier in sustained engagement in repair work.

Analytic methods
To examine how school age children—and Olivia in particular—engaged in mathematically generative play, we use methods of interaction analysis (Jordan & Henderson, 1995) to identify when children experienced trouble and worked to repair it. Trouble occurred when participants’ expectations were broken (p. 69), such as when children’s patterns and designs did not emerge as they intended. This approach allowed us to attend to both verbal and nonverbal features of interaction. Indicators of trouble include when children explicitly asked for help (e.g., “How do you make a circle?”), asked for formative feedback (e.g., “Does this look like a heart?”), expressed frustration (e.g., “Ughhhhh”), gestured inquisitively, or made trial-and-error revisions (e.g., moving pieces to make a pattern “look right”). When children persisted and made multiple repair attempts using the mathematical features of the objects, we view this as mathematically generative play, even if the trouble never was fully repaired.

Findings
Children engage in complex problem-solving during play at MOAS, and this often takes the shape of repair work. To illustrate how repair work supports children’s grappling with mathematical concepts and approximates disciplinary practices, we share a brief episode in which trouble occasioned mathematically generative construction play. Olivia worked with her mother for seven minutes to achieve Olivia’s play goal of making a heart in a 6x5 egg crate. Over the course of their activity within the play goal, Olivia articulated aspects of the heart that were important to her, which manifested as emergent problems in play. She refined her account of the emergent problem three times, with each articulation of the problem indicating an element of Olivia’s mathematics sensemaking: (1) “Does this look like a heart?” (notices shape without midline symmetry), (2) “Ohhh, it’s crooked” (focuses on asymmetry), (3) “Where’s the middle?” (focuses on egg crate as a 6x5 grid with a midline). The remainder of this section analyzes how Olivia’s question of “Where’s the middle?” launches Olivia into her ZPD as she persists in her play goal. Due to the density of gesture, which is pivotal for understanding sensemaking in constructive play, the transcript below explores 30 seconds of interaction from this last phase, beginning when she refined her problem to “Where’s the middle?” and then began to structure a
solution to the problem. Given limited space, we use only one still image from the recording for each turn, though the multimodal analysis we conducted is considerably more complex. The extra punctuation denotes speech characteristics like breaks -, elongation, (notes about gesture)), volume, and intensity.

**Transcript Excerpt**

<table>
<thead>
<tr>
<th>Talk and ((activity))</th>
<th>Screenshot</th>
<th>Analytic Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(38)</strong> Olivia: Where's the- (moves egg to slots on either side of the actual middle in the egg crate, then rests egg on actual middle)) middle?</td>
<td><img src="image" alt="Screenshot" /></td>
<td>[00:28:59.20] Olivia and Mom are in a facing-formation with joint attention on the egg crate. Olivia asks for assistance to find the practical middle in an egg crate that offers a 6x5 grid. By moving the egg, Olivia draws Mom’s attention to the absence of a midline in her current use of the crate.</td>
</tr>
<tr>
<td><strong>(39)</strong> Mom: Well that's the problem, there's- there's really no middle because this is the middle (finger runs along middle line her daughter had just touched then turns away to talk with another adult))</td>
<td><img src="image" alt="Screenshot" /></td>
<td>[00:28:59.28] Mom scaffolds attention to missing mid-line symmetry for the entire carton, then turns away, breaking the facing formation and ending joint attention to the carton as a grid.</td>
</tr>
<tr>
<td><strong>(40)</strong> Olivia: (moves body to align with side having five slots, grabs crate, then rotates body back and held crate to her original position))</td>
<td><img src="image" alt="Screenshot" /></td>
<td>[00:29:08.16] Olivia uses her whole body to investigate the practical middle of the other side of the crate.</td>
</tr>
<tr>
<td><strong>(41)</strong> Olivia: Oh maybe if we- Mommy::: (grabs Mom’s arm to get her attention, then rotates crate back to original position))</td>
<td><img src="image" alt="Screenshot" /></td>
<td>[00:29:10.14] Olivia bids urgently for Mom’s attention, pulls her body back, and restores their former facing-formation.</td>
</tr>
<tr>
<td><strong>(42)</strong> Mom: ((turns back towards Olivia))</td>
<td><img src="image" alt="Screenshot" /></td>
<td>Mom leaves adult conversation and turns her body to restore former facing-formation with her daughter.</td>
</tr>
<tr>
<td><strong>(43)</strong> Olivia: Maybe if we tu:::rn it ((rotates crate again so Mom can see the contrast))</td>
<td><img src="image" alt="Screenshot" /></td>
<td>[00:29:13.14] Olivia demonstrates her discovery to Mom by re-enacting the rotation of the crate while her mother is in their facing formation.</td>
</tr>
</tbody>
</table>
Mom: Oh yeah, good idea. Good thinking honey!

Mom recognizes Olivia’s solution to the problem of a missing mid-line (Turn 39) and praises her ingenuity with a term of endearment.

Olivia: Oh:::, no::w I can do:: it! ((takes all of the eggs out of the carton))

[00:29:27.24] Olivia clears the crate/grid and starts again to make a symmetric heart with a point/egg that falls on a true midline. Rotating the crate makes this solution possible.

Over the course of this episode of goal-based construction play, Olivia structured a new problem with new affordances for symmetry through her visible attention to aligning actual middles (i.e., midline symmetry is at the middle of six, Figure 1) with practical middles (i.e., midline symmetry is at the middle of five, Figure 2).

![Figure 1](image1.png)  ![Figure 2](image2.png)

Figure 1. No practical midline symmetry (a) with heart centered on practical middle (b). Figure 2. Practical midline symmetry (a) with heart centered on practical middle (b).

When asking “Where’s the middle?” Olivia placed her egg on the crate’s midline of symmetry (Line 38, Figure 1a), although it was not a practical middle because she had the crate rotated such that the side with six slots was parallel to her body and the egg could not rest in the center. After asking this question, Mom scaffolded Olivia’s understanding saying, “There is no middle,” while tracing the mid-line of symmetry with her hand (Line 39, Figure 1a). Importantly, Mom did not tell Olivia how to resolve the problem of no middle but rather temporarily disengaged by breaking the facing-formation (Line 39). Olivia then moved her body around the crate, making herself to parallel to the side with five egg slots. She then jointly rotated both herself and the crate back to her original position (Line 40), resulting in the crate being oriented so that its mid-line of symmetry and the practical middle for making a heart with a point became aligned (Figure 2a). Olivia then made a bid for Mom’s attention (Line 41) and, re-establishing the facing formation (Line 42), Olivia re-rotated the crate so that Mom could see the change (Line 43). Here, both Olivia and Mom demonstrated (Lines 44 and 45) that they now knew they had re-structured the problem space to align the practical and actual middles of the egg crate grid. Once Olivia removed the eggs from the crate grid, she exclaimed, “Oihhh, now I can do it!” Olivia then created her heart-with-a-point unproblematically (Figure 2b).

This episode gives a glimpse into the child’s perspective of trouble experienced in play and shows the potential for mathematical generativity in goal-based activity. First, Olivia iteratively refined her definition of the problem from not looking like a heart to eventually a problem that suggested a solution path: find or make a middle. Second, Olivia used important disciplinary practices (National Governors Association, 2010), such as (a) attending to precision as she refined her goal to be a heart with a point, (b) looking for and making use of structure as she transformed the problem space to have matching practical and actual middles, and (c) using appropriate tools strategically as she began to treat the egg crate as a grid. Third, as Olivia played, she attended to mathematical properties of the egg crate — the evenness and oddness of the two sides, the presence or absence of a true middle — and relations between them. We posit that this linking of mathematical properties facilitates deeper understanding of symmetry. Although it is not the focus of this analysis, we note that Mom effectively scaffolded Olivia, taking up her questions while also allowing her to maintain control of the decision-making process and deciding when the heart was “Done, Mommy!” (Line 63, not shown).
Conclusion and implications
This study uncovers the potential for mathematically generative play, finding that construction play with mathematically structured objects can facilitate children’s attention to mathematical properties of objects as they overcome trouble to meet practical goals. This study underscores the generativity of seemingly simple ideas (such as middles and symmetry). Future research should continue to explore this generativity, as well as examine how adults can better scaffold children’s sensemaking of key mathematical ideas. Our aim for this research program is to develop empirically-informed design principles that bring joyful disciplinary engagement into classrooms through investigating how school-age children’s play can be mathematically generative. By building an empirical basis for disciplinary enjoyment and engagement through play, this work will enable a broadening of mathematical participation in two ways. First, the mathematical ideas that children find challenging in contexts other than paper-and-pencil activities might surprise the adults who guide their thinking (Ginsburg, 2006). In other words, mathematics does not seamlessly transfer between scales and modalities (Hall, Ma, & Nemirovsky, 2015), so intentional scaffolds must be designed as we create wider and deeper ecosystems for mathematics education. Second, children’s competent informal mathematical thinking is often not recognized by adults (Ginsburg, 2006). By capturing examples of children’s productive, informal mathematical thinking, this research can facilitate mathematics educators’ recognition of swaths of competence. By underscoring the inherent challenge of seemingly simple ideas such as middles and symmetry, adults can be better attuned to helping children make meaning of key mathematical ideas.

References


Teaching with a Fully Digital, Year-long Math Program: Learning Science Futures on the Front Line

Jeremy Roschelle, Digital Promise, jroschelle@digitalpromise.org
Phillip Herman, McREL, pherman@mcrel.org
Karen Bumgardner, McREL, kbumgardner@mcrel.org
Nicole Shechtman, SRI International, nicole.shechtman@sri.com
Mingyu Feng, SRI International, mingyu.feng@sri.com

Abstract: In the context of a large-scale randomized controlled trial, our team investigated “front line” teaching issues as schools implemented a fully digital, blended learning curriculum in mathematics. This paper focuses on observations of instruction within schools that were assigned to use the new digital resources. Compared to a business-as-usual control group, classroom activity and teaching practices changed in the treatment group. Observers, who were blind to student achievement outcomes, found two overall patterns in treatment classrooms across five categories of observations. Later quantitative analysis indeed found the “high” and “low” patterns could account for some of the variance in achievement outcomes within the treatment condition. We explore observed patterns in terms of existing learning science theory and suggest areas where further development of the learning sciences may be needed and how learning sciences can contribute to improvement of digital, blended learning environments.

Introduction

As indicated in this year’s conference theme, AI and automation are changing the nature of classrooms as workplaces for teaching and learning. As these changes occur, new complexities arise for learning scientists who study classrooms. As the work of teaching and learning becomes distributed across teacher and technology, learning scientists may need to change or refine understandings of effective teaching and learning processes. The conference theme draws attention to the “imperative to guide commercial development,” as well as the need to understand different cultural and educational contexts. Our research investigated a commercially-available digital curriculum and we worked closely with the product team to understand the lessons of the study for improvement. We also worked in schools in West Virginia, a state with a distinctive regional culture. This paper discusses how the learning sciences may need to evolve in order to be responsive to the vision in the conference theme.

We conducted a randomized controlled trial (RCT) aimed at measuring the efficacy of a new digital mathematics curriculum, engaging 46 schools and approximately 2000 students in our research. Overall, we found that the vision and challenges described in the conference program to ring true: AI is changing the classroom workplace. We observed the classrooms in both the treatment condition (TC), which was using the new digital materials and blended learning approach, and in the control condition (CC), which was using “business-as-usual” materials and approaches. As we will later describe, we observed broad differences in the structure of classrooms, such as the predominance of instructor-led mathematics teaching (CC) versus individual student work at computers (TC). For students, the “work of learning” in TC classrooms was also different. For example, TC classrooms emphasized independent learning strategies but CC classrooms did not. For teachers, the change in the role was quite extensive. For example, teachers were no longer the main providers of instruction. Further, teachers in TC classrooms were more likely to use data during class to make instructional decisions; TC teachers spent much time intervening with particular students based on data reports (Singleton et al., 2018).

This paper’s investigation of this new AI-rich classroom workplace focuses on classroom observations but leverages data from the larger RCT as well. The RCT engaged a large number of schools and our team of observers collected data in all 23 TC schools and with all 38 TC teachers. Thus, we have the opportunity to look at systematic patterns of teaching and learning that were emergent across many teachers and schools, whereas many learning science studies only look within a handful of classrooms. Another advantage of the RCT context is that we collected achievement outcome data for all schools, and can control for prior achievement in our analyses. This allows us to ask: controlling for prior knowledge, do the systematic patterns we observed in different TC classrooms predict differential student outcomes in those classrooms? Thus, we can examine whether there is evidence that the observed patterns might be consequential.

To achieve the kinds of positive impacts envisioned in the conference program, learning scientists need to know if their theories are a good match to what goes on in new digital, blended learning classroom environments. We argue that the identification of systematic patterns that are consequential can be a good guide to where the learning sciences, if further developed, could have stronger impacts with regard to new AI-based...
teaching and learning workplaces, guiding commercial products, and working within specific cultural settings. Within the broad frame of learner-, knowledge-, assessment- and community-centered classrooms drawn from the seminal *How People Learn* (Bransford, Brown & Cocking, 2000), we use our findings to suggest implications for ways in which the learning sciences may be fruitfully developed.

**Context: The Learning Sciences as co-evolving with technologies**

The learning sciences have always closely connected new understandings of how people learn (HPL) to emerging new technologies for and approaches to learning (Bransford, Brophy & Williams, 2000). For example, earlier advances in technology made new representations of mathematics possible, such as dynamically linked multiple representations. Learning scientists studied how students make sense of mathematics with linked representations (Roschelle, Noss, Blikstein & Jackiw, 2017). Likewise, a long-standing program of artificial intelligence in education made it possible to trace (assess) student knowledge and give targeted feedback, and researchers studied how learner- and assessment-centered AI approaches could improve learning (Luckin, Holmes, Griffiths & Fourcier, 2016). Many earlier learning science studies only examined short curricular unit, because this is what it was feasible to field across many classrooms. Further, earlier studies often examined only a few schools or classrooms, because getting the necessary technology in place was often hard.

Now technological platforms make it feasible to deploy techniques like multiple representations and AI in a curricular resource that spans a full classroom year. Further, the collection and rapid use of student data has become easier, and it is possible to provide teachers with dashboards and reports to guide their work in real time. New approaches such as “blended learning” are becoming popular among educators. In blended learning, it is expected that teachers and technologies will each have a complementary role in the overall instructional program (Means, Toyama, Murphy & Bakia, 2013). Importantly, these infrastructures and approaches have become sufficiently commonplace that they can be studied not just in special research-partner schools, but in a sample of schools recruited from a whole state. Programs that are year-long and could scale state-wide could have big impacts. For learning sciences to play a role in understanding these impacts, it may have to adjust its focus. In this paper, we will look at that evolution in terms of the HPL framework of a learner-, knowledge-, assessment- and community-centered classroom.

**The math curriculum impact study**

This study, funded by the Institute of Educational Sciences in the US, was intended to investigate the efficacy of a year-long, digital, blended mathematics curriculum with a strong AI component. The main hypothesis was that grade 5 students in schools that implemented the new mathematics curriculum for a full year would have higher mathematics achievement at year end than in schools in a business-as-usual control condition.

**Intervention: Reasoning Mind**

In the TC, schools were asked to use Reasoning Mind’s grade 5 core curriculum (hereafter, “RM”) as their main instructional resource. With regard to being knowledge-centered, RM’s instructional approach (Khachatryan et al, 2014) is closely modeled on an exemplary international approach and seeks to build complementary facets of mathematical ability: fluency with calculations and deep understanding of foundational concepts. It also has a strong problem-solving component, with three levels of progressively harder problems and a “smarter solving” module. With regard to being assessment-centered, RM collects copious data as students do mathematical work online and continuously monitors student progress. These data are used to ensure the system is learner-centered. The system adapts its instruction, the difficulty of problems, and the pace through the materials based on AI techniques, so that instruction is personalized for each learner. Further, teachers get useful reports that guide their work with specific students (or groups of student). Teachers can assign special assessments to follow up and see if their interventions with students paid off or more support is needed. Another important aspect of the learner-centered approach in RM that aligns with HPL is the focus on metacognition; RM’s pedagogical approach seeks to develop independent learning strategies, such as students keeping good notebooks and using them when they get stuck, rather than always asking a teacher for help. RM also is notably community-centered. The program includes whole class incentives to motivate students. RM builds a strong classroom mathematical culture in part by introducing a “Genie” character to whom students relate and who establishes norms for a mathematics learning. RM envisions and supports a classroom community where students learn individually, but also where they support each other and celebrate successes together, and where teachers have time to care for the needs of individual students. Reasoning Mind was also an attractive intervention to study because it had good prior results (Roschelle, Bhanot, Patton & Gallagher, 2015) and a strong capability to achieve high quality implementation in many schools at once through the role of Implementation Coordinators (Roschelle, Gaudino & Darling, 2016). Previous studies had also found high levels of student engagement in classrooms using RM (Ocumpaugh et al 2013).
Setting and sample: West Virginia schools
We conducted this study in West Virginia (WV), a state shaped by its geographical setting amongst the Appalachian Mountains; mountains and rolling hills define the region. Our WV-based McREL team expressed that their region has a love of place, community, and family, and in our initial contact with schools, we felt we could see these attributes reflected in the classroom. WV has low population density and the median household income of $42,000 is considerably lower than the national median of $56,000 (Frohlich, Sauter & Stebbins, 2016). Throughout the US, low family income and lower mathematics achievement are correlated. WV has been a leader in putting strong computing facilities in its schools and connecting schools to the Internet with high bandwidth. Since 2011, access to wired connections has improved from approximately 45% to 91% of West Virginians (Broadbandnow, 2017). The WV State Board of Education has adopted as its goals to “provide a high-quality learning system that (a) encourages a lifelong pursuit of knowledge and skills, (b) promotes a culture of responsibility, personal well-being and community engagement and (c) responds to workforce and economic demands.” Just prior to this study, WV also adopted curriculum standards in mathematics that set high expectations for all students, and the teachers we worked with showed strong commitment and effort towards increased mathematics achievement. RM already had an implementation in a few WV schools that was going well, which made recruiting easier. We recruited over 50 schools to participate in the two-year randomized controlled trial from districts spanning the state, and although a few dropped out for various reasons, 46 schools remain in the final data sample we analyzed.

Research design
We planned and conducted a randomized control trial. Schools were matched in pairs that had similar prior math scores and geographic locations, and then a coin was flipped for each pair. Schools assigned to the TC were trained and supported to use RM for two years: a “warm up” year in which the teachers learned the new pedagogical approach and a “measurement year” in which we collected student prior and end-of-year achievement data. Schools assigned to the CC continued with business-as-usual materials and teaching approaches for 5th grade, but as an incentive, they were offered a different RM product for use in grade 2. These students would not reach grade 5 until the study was over. No CC teachers taught both grade 5 and grade 2, to avoid contamination.

Measures
The study collected a very rich array of measures, including teacher interviews and surveys and RM system data. However, in the scope of this paper, we focus on only two measures, a standardized test and observations. We used the required statewide assessment, the WVGSA, for both end-of-year mathematics achievement in grade 5 and as a prior achievement covariate in grade 4. This assessment was designed to be adaptive, to align with the state’s curriculum framework, and to measure problem solving and not just procedural fluency.

A team at McREL designed an observational measure. Designing this measure was a challenge, because we wanted a measure that would work in both the TC and CC and the nature of what could be observed in these settings turned out to be quite different. A manuscript under development will describe in more detail how the measure was designed and refined through different phases of pilot testing in schools (Herman & Bumgardner, in preparation). In the course of the refinement process, the McREL team revised the instrument and their training until sufficient interrater reliability (> 80%) was achieved.

We report only on observations from the measurement year, which used an observational instrument that was agreed upon by all the partners in the study. In the instrument, an observer in the classroom typed running record field notes, guided by a framework with five areas in of observation: (a) on task behavior, (b) motivational routines, (c) independent learning strategies, (d) use of data, and the (e) quality of mathematical discussion among teachers and students. After making field notes, each observer rated these five areas on a 1 to 3 scale where “1” (low or not present or rarely) to “3” (high or frequently), using a rubric that set criteria for the scale levels.

Data
We obtained statewide scores for both grade 4 and grade 5 for students in 46 schools, 23 TC and 23 CC. We conducted a total of 53 observations in 38 TC classrooms and 15 CC classrooms. We included all the TC classrooms because of the greater interest in these, but only a sample of CC classrooms due to limited budget. For each observation, we collected a running record plus ratings for that classroom.

Analysis plan
For the main impact analysis, the SRI research team set up a two-level hierarchical linear model to account for the clustering of students within schools, using the grade 5 student scores as an outcome variable and the grade 4 scores as a co-variante. In later models, ratings from observations were added as potential mediating variables that
might account for additional variance. The main impact analysis is the subject of a forthcoming journal submission (Shechtman et al., 2018) and is not reported in detail here. For the analysis of observations, the McREL team looked both at the contrast between observations in the CC and TC and also variation within only the TC. The McREL team analyzed its observational data for potential variations among classrooms in the TC that might be systematic across a set of teachers. They did so without awareness of which teachers or schools had achieved higher or lower mathematics achievement outcomes. Independently, the McREL team developed its own sense of “low” vs. “high” implementations meeting as a team to review its ratings (through the observation process) of the schools it observed, and then analyzed field notes to see if common themes emerged. For further details on the quantitative rating, see (Herman & Bumgardner, 2018).

Findings
We discuss the main impact findings briefly, for context. Full presentation of findings will be in (Shechtman et al., 2018). We then also consider the contrast between TC and CC. We focus thereafter on the findings about variation within the TC, and look for systematic ways in which TC classrooms varied and examine evidence as to whether those variations were plausibly related to student achievement outcomes.

Impact findings
In contrast to our hypothesis, the data did not reveal a measurable difference in mathematics achievement between TC and CC schools on the WVGSA test. Although unexpected, this does not mean that the TC was bad for students, indeed the student outcomes did not differ significantly between groups.

We also found considerable variation within the TC. In Figure 1, we illustrate this graphically by drawing a bar for each school where the height represents the average mathematics score at the end of grade 5 for that school. The bars are ranked by score and filled by condition. This shows that some schools in TC had some of the highest end-of-year math scores, but other schools in the TC were among the lowest scoring schools. This distribution led us to wonder about differences between the schools at each end of the distribution, for example, differences in how they did the work of teaching and learning.

Contrast findings (treatment vs. control)
Our observers found striking differences between TC and CC classrooms. In general, observed compliance to assigned condition was high: TC classrooms were observed to be implementing RM as the central instructional resource and CC classrooms were not. Control classrooms looked reasonably traditional. For example, the teacher often provided instruction from the front of the room and students often worked on math problems individually or in small groups as a teacher (and often additional instructional aides) walked around providing help. Technology was available, but not used frequently. In TC (RM) classrooms, students sat down at computers at the beginning of their mathematics lesson and began working individually; most students this way most of the time. Teachers typically had a station in a corner of the room with a computer that provided reports on student work. The teachers called individual students (or small groups) to their station and worked on targeted issues. Aside from these interventions, there was not as much small group work as in CC classrooms.

With regard to the five scales in the observation instrument, the observers did not find any statistically significant differences between conditions in the degree to which students were on task nor in the observed supports for motivation and engagement. There were two statistically significant difference that favored the TC: as expected, there was more data use (92% vs. 8%) and more emphasis on independent learning strategies (88% vs. 12%). The observers also rated control classroom as higher on the quality of mathematics instruction scale. However, they also noted that it was harder to make relevant observations on this scale in the TC classrooms – for example, more of quality of instruction was mediated by technology and was hard to observe.
Variation within treatment findings

High rated classrooms
Nine of the thirty-eight treatment classrooms (24%) received overall ratings of 3 across all subscales. The instruction strategies, techniques, and procedures observed across each of these nine classrooms were quite similar, with recurring themes emerging. High scoring treatment classrooms had teachers who demonstrated comfort and control in managing student behaviors in their classrooms, structuring their classes and lessons in a manner conducive to student engagement and learning. Students in such classrooms regularly demonstrated familiarity with behavioral expectations and performance objectives, as well as standard classroom procedures. In many treatment classes observed, students entered the classrooms at the beginning of the period, obtained laptop computers, and signed in to RM to begin their coursework without prompting from instructors.

Almost all highly-rated treatment classrooms concluded the lessons with a review of students’ performance for the day, with teachers highlighting mathematical achievements at both the whole-class and individual level. Students in highly-rated treatment classrooms demonstrated apparent investment in their mathematical performance—for example, in their commitment to engaging with the RM system. All highly-rated treatment teachers were observed employing adaptive learning strategies and techniques that aligned with RM. For example, they focused on independent learning strategies, by asking students to use available resources to resolve mathematical difficulties before calling a teacher over. Teachers in these classrooms also frequently used formative performance data in real-time, using it both to motivate students and to select groups of students for one-on-one interventions. With regard to motivation, teachers in the high rated classrooms, tended to exhibit community-centeredness by featuring the whole classes daily performance statistics before discussing any individual students. In the one-on-one interventions, the mathematics talk in these classrooms more often involved students in doing significant amounts of mathematical work; the teacher didn’t do the math for students. However, overall, it was hard to observe mathematical knowledge building, in part because teachers often directed students to do work on the computers; further, the knowledge-building work that was available in discourse was like an “intervention” than a longer-term process of developing understanding.

Low rated classrooms
McREL observers rated six of thirty-eight (16%) of classrooms as lower quality across all categories of observation. A wider variety of instructional strategies, policies, and procedures were observed across lower-rated compared to higher-rated treatment classrooms, though several patterns were consistent. Perhaps the most overt trend observed across lower-rated treatment classrooms was the extent to which many of the teachers appeared to struggle in implementing effective classroom management. Some students spent extended periods of time disengaged from mathematical material in these classrooms. These teachers demonstrated an ability to recognize off-task behavior, but demonstrated difficulty in sufficiently addressing it.

Lower-rated treatment classrooms also differed from one another as far as the extent to which teachers established RM-related routines and expectations as well as the extent to which objectives were evident and communicated to students. Like higher-rated treatment classrooms, several teachers in lower-rated classrooms clearly communicated and referenced established procedures and ensured that students understood what was expected of them. In other scenarios, teachers were not observed making significant effort to motivate students to use RM as intended. Teachers from three classrooms were not observed incentivizing or encouraging engagement with mathematical material to a significant degree, and, thus, students in these classrooms appeared predominantly ambivalent as to the extent to which they accomplished RM objectives. Additionally, none of the teachers in lower-rated treatment classrooms were observed implementing strategies or practices to facilitate student autonomy or independent learning strategies. Across each of the lower-rated treatment classrooms, students were reliant on course instructors to make significant progress through the RM curriculum, with several students exhibiting an inability to engage in any mathematical work independently. Further, in some classrooms, teachers actively inhibited students from independently completing assignments—emphasizing more teacher control in the mathematical work of the classroom. In lower-rated treatment classrooms where teachers allowed students to collaborate with classmates, students nevertheless appeared to have trouble engaging with mathematical content without assistance from instructors. None of the teachers in these classrooms were observed making references to resources that students could use for math learning without involving the teacher. On occasion, a student or small group of students appeared to consult the hints provided in RM or refer to the RM library, but these behaviors were rare. For the most part, teachers of low-rated treatment classrooms were not observed making frequent use of data to inform instruction. Those teachers who used data did so in a more supplemental manner.
Exploratory model
To explore whether the high and low patterns identified by McREL’s observation team might relate to student outcomes, we conducted an additional analysis. For this analysis, we included only students of TC teachers in either the high (n=10) or low (n=6) pattern classrooms. We conducted a 2-way ANOVA in which the outcome variable was assessment score, and the factors were (1) Year (Grade 4, Grade 5), (2) McREL Group (Low, High), and (3) the interaction term (Year x McREL Group). Means are shown in Figure 2. We found that both main factors were significant. As would be expected, WVGSA scores were higher in Grade 5 than in Grade 4, F(1,748)=12.95, p <.001 (while reported on the same scale, the tests were different and aligned with respective grades). There was also a main effect of McREL Group, such that students of teachers in High classrooms had higher assessment scores than did students of teachers in Low classrooms, F(1,748)=8.40, p <.01. The interaction term was not significant, F(1,748)=.43, p = .51, n.s. In everyday terms, this means there was no closing nor expanding of the achievement gap.

Discussion
Overall, the work of teaching and learning was quite different with the digital, blended learning approach – treatment classrooms were quite different from control classrooms. Further, based on observations, we were able to identify systematic “high” and “low” patterns within the treatment schools. We explored the importance of these patterns in a quantitative model that included student prior achievement and student outcomes. The observed patterns appear to be linked to classroom mean prior achievement, which raises equity issues. Overall, only 16% (6 of 38) of the observed classrooms fit the low pattern. While examining these classrooms is useful in looking for improvements, these classrooms are not representative. We emphasize exploratory investigation of these classrooms, and do not use this small sample to reach generalizable conclusions. We frame our discussion in terms of “uptake” – the uptake of unique RM features and possibilities was different across the two classroom groups.

Learner-centered
In high functioning TC classrooms, there was more uptake of learner-centered opportunities. For example, high functioning classrooms were observed to place an emphasis on independent learning strategies, but low functioning classrooms did not. Likewise, on the quality of mathematics instruction scale, we found that math talk in the higher functioning classroom gave the responsibility for doing mathematical work to the students, whereas in lower functioning classrooms, teachers did more mathematical work. The suggests that not all classrooms take advantage of the opportunities for learner-centered instruction equally. However, one caution is that the observed low group also had lower prior mathematics achievement scores; it could be that some students (who are about 10-11 years old in grade 5) are not ready for independent learning strategies in mathematics and may benefit from a more teacher-centered approach. Likewise, teachers may have a stronger repertoire for engaging students with higher mathematics achievement in doing mathematical work, but may default to doing more of the mathematical work for their lower-performing students. In a traditional classroom, this may be less evident, because there may be enough mathematical knowledge in the classroom overall for the teacher to sustain a high quality of mathematical discourse. The learning sciences may need to elaborate how teachers could enact learner-centered instruction when they are working one-on-one with a large group of students who are coming in with low existing knowledge in mathematics and really struggling.

Assessment-centered
Overall TC classrooms used real-time data reports to make instructional decisions, whereas CC classrooms did not. Moreover, within TC classrooms that were observed to be lower functioning, there was much less use of real-time data reports. Again, we urge some care in interpretation. It could be that implementation coordinators should help teachers to make better use of the data reports. But it also could be that certain factors have to be in place before teachers can sensibly use data reports in real time. For example, if students do not stay on task during individual work at computers, it may not make sense for teachers to be looking at data reports during classroom time. Likewise, if students are uniformly struggling with the mathematics (the observed-low group had weaker prior math achievement), it may make less sense to teachers to work with individual students. The learning sciences could help us understand better how to leverage a more assessment-centered teaching structure with classrooms that are more or less ready to engage in grade-level mathematics.

Community-centered
As learning communities, the observed-low classrooms were more chaotic, with many behavioral problems. It is unclear why the 16% of classrooms in the “low” group had lower engagement. We are reluctant to attribute it to the students alone, because control classrooms with low achieving students did not have the same level of

© ISLS
behavioral problems. It could be that the experience of digital and blended learning was less satisfying to some groups of students, and the behavioral problems emerged from their frustration and confusion. However, this was not observed in prior studies of student engagement with RM (e.g., Ocumpaugh et al, 2013). Our team of observers sometimes wondered whether the mix of instructional activities in the TC classrooms had too much of an emphasis on individual time at a computer, and whether the classroom community might fare better with social activities like small group work and full classroom discussions. We particularly wondered whether the rather quiet and individualistic classrooms in the blended learning condition may not have fully utilized the community-centeredness otherwise evident in the WV classrooms that we visited. Yet, in the observed-high group, we were able to see many positive community-centered aspects of classrooms, such as peers helping fellow students and new norms being established through the character of the Genie. We noted that teachers spent lots of time working one-on-one with students, which can be good for strengthening relationships.

Knowledge-centered

It was hard to observe how knowledge building worked in the new, digital learning blended environment. This may be a weakness of observational methods. With regard to the method, classic learning science “knowledge building” environments make more use of collaborative, social and full-classroom learning – the TC classrooms were less collaborative and social and thus knowledge building was less public. Methodologically, we were able to observe strengths of the RM instructional materials as we watched individual students use them, for example, the recommendation to read “Genie solutions” to understand problem solving processes, which related to the well-known self-explanation effect (Van Lehn, Jones & Chi, 1992). The RM materials also give strong conceptual presentations to students, and these may be stronger that many teachers would typically achieve in their own presentations of concepts. Yet, although there is more time and space in RM classrooms for teachers to work with individual students on their conceptual understanding of mathematics, we observed variability in the quality of the mathematical discourse in TC classrooms – many of the conversations we observed were more procedural than knowledge-building oriented. Overall, a challenge for the learning sciences is to come to a better understanding of how to measure knowledge-building in environments like those we observed in TC classrooms, which less available public knowledge-building discourse.

Equity

Overall our analysis revealed a potential equity issue. We observed less uptake of the capabilities of the digital, blended curriculum and more behavioral management problems in a cluster of classrooms, and then later found the cluster had lower mean prior achievement. Conversely, when our observers noted a cluster with uniformly high quality of implementation, we found that the prior achievement scores in those classrooms were higher. Causality cannot be determined from this analysis. Either (a) while the digital, blended approach is appropriate for classrooms with lower mean prior achievement, specific additional support for implementation is needed or (b) the adaptive, blended learning approach may have been less appropriate for classrooms with lower prior mean achievement, running into classroom problems despite good implementation support. Overall, we remind readers that there is a distribution of low-to-high achieving students in every classroom, so it cannot be inferred from this analysis whether the approach has differential benefits for individual students who have higher or lower achievement (e.g., there were some students with lower achievement in the classrooms that had higher mean achievement). The only relationship we explored was between classrooms that were observed to be making lower use of the HPL-related features of RM and classrooms that as a whole had lower mean prior achievement.

Conclusion

The ICLS 2018 conference program envisions a future in which the learning sciences helps schools to make sense of new teaching and learning approaches, including those with strong AI components, and guides improvement in the quality of products. The MCIS project provided an opportunity to investigate a future-oriented learning environment, with a full-year digital curriculum that incorporated AI features and a blended learning instructional approach. We found that schools were able to implement this curriculum throughout a state. Relative to classrooms in the control condition, there was a strong contrast in how the new classrooms functioned as workplaces for teaching and learning. We also looked systematically across implementing classrooms for systematic patterns that might explain variation in classroom outcomes. We found a pattern in which a group of classrooms with lower uptake of the HPL-related features of the new technological approach also were classrooms with lower prior mathematics achievement. This points us to one way in which Learning Scientists could help product developers – by examining the differential uptake of research-aligned features and examining relationships to equity factors, like low prior achievement. As our data does not reveal causality, we would encourage future researchers to explore why uptake of HPL-aligned features was lower in some classrooms.
We also believe that the Learning Sciences itself needs to change in order to become more relevant. We need better methods for making sense of knowledge building activities when distributed and so thoroughly mediated by technology that there is more individual and less “community” knowledge building time. The learning sciences could say more about how to make sense of the kinds of observations we made of lower functioning classrooms: were students ready for an emphasis on independent learning? How can teachers have good mathematical conversations in one-on-one settings, where they may less variety of students’ ideas to draw on than in full class discussions? How do teachers decide when to make use of student progress reports, and are there circumstances in which using such reports is more or less useful? Overall, we suspect the learning sciences could make stronger contributions as envisioned in the conference program with greater attention to systematically describing variability at scale and helping to uncover aspects of variability (and equity) which are consequential for learning outcomes. In this way, we foresee more relevance for the learning sciences by aligning with improvement sciences and through a focus on measuring and addressing undesirable variability across settings.

References


Acknowledgements

This material is based upon work supported by the Institute of Educational Sciences (IES) of U.S. Department of Education under Grant Number R305A130400. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of IES. We thank the West Virginia Department of Education for providing access to the data. We also thank the students, teachers, and schools who participated in the study as well as the project staff from Digital Promise, SRI International, McREL, and Reasoning Mind.

ICLS 2018 Proceedings 639 © ISLS
Interlacing Gaze and Actions to Explain the Debugging Process

Kshitij Sharma, Norwegian University of Science and Technology, kshitij.sharma@ntnu.no
Katerina Mangaroska, Norwegian University of Science and Technology, katerina.mangaroska@ntnu.no
Michail Giannakos, Norwegian University of Science and Technology, michailg@ntnu.no
Pierre Dillenbourg, Ecole Polytechnique Fédérale de Lausanne, pierre.dillenbourg@epfl.ch

Abstract: Debugging is an indispensable skill of successful programmers. As such, teacher’s should not overlook it when teaching programming. The main aim of the study is to use gaze data combined with measures at different temporal granularities to show how these measures are related to the outcomes (in particular debugging success as a learning by doing outcome) students have at the end of the debugging task. The results delineate that combining gaze data with actions (reading, writing, scrolling) and unit tasks (main method and JUnit test) gives new insights to further understand the cognitive actions in debugging a program. Moreover, this study also focuses on discovering debugging patterns of successful students in order to improve the design of learning activities to teach students how to debug. Finally, with the analysis, the authors have shown an automatic way of detecting successful action-gaze patterns.

Keywords: Computer science education, Eye-tracking, Time-scales, Program debugging, Programming.

Introduction
One of the most essential 21st century skills is programming. We have witnessed the influence of programming in disciplines like engineering, biology, chemistry, etc. Therefore, many engineering programs require students to take introductory programming classes. However, even today there is still no solid retention in programming courses. In addition, (McCracken et al., 2001) assessed the skills and competences of first year programming students concluding that many freshman cannot write correct code. One of the reasons behind this issue is the difficulties and challenges students face while learning to program. An example of this is debugging. Besides being vital, learning to debug is difficult and requires application of many skills simultaneously that majority of freshman do not have at the beginning (Perkins & Martin, 1986). However, debugging is an indispensable skill of successful programmers and cannot be neglect it when teaching programming. Moreover, there are no best practices or pedagogies available for programming educators to teach “how to debug” (Fitzgerald et al., 2010).

Another reason is lack of computational thinking skills. Computational thinking is a fundamental skill for everyone, not just for programmers (Wing, 2006). It is way humans solve problems, by thinking at multiple levels of abstraction. It also means applying heuristic reasoning in presence of uncertainties when solving complex problems. As such, computational thinking should be part of the skill set of future programmers. Consequently, if programming educators want to improve the curriculum and the teaching, they need to understand how students solve problems, what resources they use, what strategies they follow, and what motivates them to continue to learn programming. Eye-tracking can help them to get closer to the students’ way of learning programming and observe actions that cannot be captured by observations or through system log data.

Eye-tracking allows researchers to get attentional data for users while they perform tasks at a time scale that has more information content than other measures like event logs, dialogues, or gestures (Sharma, Jermann, Nüssli, & Dillenbourg, 2012). (Newell, 1994) proposed time scales and human action bands to describe behavior at various levels. For example, the duration of gaze fixations is usually 100 milliseconds, placing it at the lower cognitive band. Lower cognitive bands are the primitive actions (Newell’s Deliberate Acts) that Newell believed if combined the right way at this level can lead to understanding the development of higher level constructs, such as expertise. This has been supported in studies like Gluck’s study of instructional leverage with an algebra tutor (Gluck, 1999) or Ackerman’s study of processing parallel threads in air traffic controller task (Ackerman, 1994). At the higher end, cognitive bands require complex actions (e.g. reading or gestures). Since there is a time gap between significant educational outcomes (i.e. tens of hours to achieve) and effects measured in tens of milliseconds, (Anderson, 2002) identifies cognitive modeling as bridging across the behavioral bands by taking the lower level bands into account. The main aim of this study is to use gaze data to define measures at different temporal granularities, and show how these measures are related to the outcomes (in particular debugging success as a learning by doing outcome) students have at the end of the debugging task.
addition, we will use screen recording from the eye-trackers to disambiguate the actions into reading, writing and scrolling episode, and find patterns that successful students follow. These patterns could later indicate opportunities for instructional leverage. In this contribution the study was driven by the following research question: “**How the interplay between the gaze patterns and actions explains the debugging success?**”

The rest of the paper is organized as follows: the second section is the related work for the present study. The third section describes the research methodology. The fourth section describes the various variables used in the experiment. The fifth section presents the analysis results. Finally, the sixth section discusses the results and concludes the paper.

**Related work**

**Eye-tracking and education**

Studying successful and unsuccessful programmers provide important insights for improving teaching because it allows to understand how programmers learn and what misconceptions they might have. (Vessey, 1985) found out that expert debuggers were following a breadth first approach, trying to understand a program and build mental representations, while novices were following depth first approach, focusing on finding errors rather than understanding the program. Consequently, if programming educators have such insights, it might help them to design correct models “how to teach novices”. Adapting to the technological advancements, gaze data has already proven its value in understanding how students learn to program and debug (Sharma, Jermann, Nüssli, & Dillenbourg, 2013). Eye-tracking helps researchers to gain insights into user behavior (e.g. how the user process information or interacts with visual information) that cannot be captured verbally, via other more ordinal user-data (e.g. click-stream) or with observations (Cooke, 2005). Using eye-tracking to investigate differences between experts and novices, but also between “good” and “poor” novices, could give new insights that could be used to develop instructions for a specific group of learners. In addition, researchers could also gain further understanding what is happening in the stages of understanding the program, testing the program, or locating the errors in the program.

**Eye-tracking and debugging**

Most results show that eye-tracking allows researchers to get users’ attention, while they perform tasks to explain various constructs like contextual expertise, task-based performance, and task complexity (Harbluk, Noy, Trbovich, & Eizenman, 2007; Kaller, Rahm, Bolkienius, & Unterrainer, 2009; Reingold, Charness, Pomplun, & Stemple, 2001). Furthermore, “debugging is a skill that does not immediately follow from the ability to write code. Rather…it must be taught” (p. 208) (Kessler & Anderson, 1986). In addition, a comprehensive review of debugging in education has been performed; however, little has been done after it to gain new insights (McCauley et al., 2008).

Previous studies (Bednarik, 2012; Bednarik & Tukiainen, 2004, 2008) showed a clear relation between gaze patterns and task-based performance in debugging, underlying the importance of using eye-tracking to understand students’ patterns when running into difficulties while working on problem-solving tasks (e.g. debugging code). Thus, in order to avoid programming errors that result from misconceptions, programming educators should not ignore incorrect mental models students might have, but must understand their constructs that provide insights how to design curriculums to teach programming/ debugging techniques (McCauley et al., 2008). For example, in (Sharif, Falcone, & Maletic, 2012) authors compared the first scan time (the time takes by the participants to read the code for the first time) against the different levels of debugging success. The results showed that successful debuggers had a significantly lower first scan time than the less successful ones. In terms of gaze behavior, this study showed that successful debuggers had a more vertical gaze than those who perform less successfully during the debugging task. Moreover, (Stein & Brennan, 2004) in a study considering a collaborative debugging task, showed that a gaze-contingent condition (to find a bug after viewing gaze trace as a visual cue) induced a better way for finding bugs in a given code. Like-wise, (McDowell, Werner, Bullock, & Fernald, 2006) came to the same conclusion years before, unveiling that students working in pairs are more successful, more confident, and more likely to continue studying programming. One common drawback of most of the previous studies (to the best of our knowledge) concerning debugging, is that the task was limited to “*find the bug*” and not actually to “*fix the bug*”. This limits these studies to a mere comprehension task. In the present study we asked the students to “*find and fix the bug*”. Thus, our study captures more in-depth data (both in terms of actions and gaze) in a debugging task.
Methodology

The debugging activity
The authors designed and implemented a debugging activity in conjunction with the partners from the École Polytechnique Fédérale de Lausanne University. The main task assigned to the participants was debugging a code that contained consistency issues implemented as unit tests. The consistencies that were absent from the original version of the code provided to the participants are the following: 1) Gender consistency: the mother should be a female and the father should be a male; 2) Child-parent consistency: if Jens is the child of Merit, Merit should be the mother of Jens; and vice-versa; 3) The removal of a child-parent relationship from either a parent or a child should also apply to the whole family; 4) Adoption consistency: the child-parent (addition and removal) and the gender consistencies should be maintained in the case of an adoption.

Participants
During the spring 2017, an experiment was conducted at a contrived computer lab setting at École Polytechnique Fédérale de Lausanne University with 40 computer science majors (12 females and 28 males) in their third semester. The mean age of the participants was 19.5 years (Std. Dev. = 1.65 years). In the previous semester, all of the participants had taken a Java course, where they were predominantly using Eclipse as Integrated Development Environment (IDE). Moreover, they were also familiar with the built-in debugging tool provided by Eclipse.

Procedure
Upon arrival in the laboratory, the participants signed an informed consent form. After this and prior to the debugging task, each participant had to pass an automatic eye-tracking calibration routine to accommodate the eye tracker's parameters to each participant’s eyes to ensure accuracy in tracking the gaze. Their gaze during the debugging task was recorded using an SMI RED 250 eye-tracker at 250 Hz. Next, the participants were asked to perform a pre-task, which required removing 90 errors from a skeleton code within ten minutes. After this task, the participants were given 40 minutes to solve five debugging tasks presented as a part of the main method of the main class of a 100 lines of Java code. The code for the main debugging task contained no syntactic errors, and the participants were notified about this fact. For their participation in the experiment, the participants were rewarded with an equivalent of USD 30.

Variables

Debugging success
For the debugging task, there were ten unit tests prepared by the instructor. These unit tests were about the consistency of the parent-child relationship (see Subsection “The debugging procedure”). To limit the debugging to one of the panels of the Eclipse IDE, the researchers introduced few bugs in otherwise complete code that would make the code fail all ten unit tests. In order to pass all of the unit tests, the students were required to solve the debugging exercises in a particular order. For example, the gender consistency needed to be solved before the child-parent consistency. Participants were given 40 minutes to complete the task. At the end of the 40 minutes, they were told to stop, and the number of unit tests passed at that point of time was taken to be the measure of the “debugging success”. Hereafter we will refer to “debugging success” as “success”.

Unit tasks: JUnit tests and main method
In order to complete the debugging task, the participants were testing their solutions multiple times. They could do this in two different ways: 1) by running the set of JUnit tests, or 2) by executing the main method of the program. We divided the whole debugging session into two types of episodes based on the unit tasks performed by the participants namely, JUnit and main.

Actions: Reading, writing, and scrolling episodes
Participants’ screen was recorded during the debugging task (at 10 frames per second). We computed framewise image difference, in order to detect reading, writing, or scrolling episodes. First, we assigned, to each frame, an action flag denoting whether it was a reading, writing, or scrolling frame. For doing so, we used two thresholds for the number of pixels changed across two frames. If there were no pixels changed, we assigned the frame a reading flag; if the change was corresponding to a range of 0.4 -- 0.8 characters changed over two frames, we assigned the frame a writing flag; for a change corresponding to more than 10 characters per frame we assigned the frame a scrolling flag. For obtaining 0.4 pixels to 0.8 pixels, for the writing labels, we used the following
logic. The average person’s typing speed is 40 words per minute, the average length of each word in JAVA could be taken as 6 characters. This translates to 240 characters per minute, or 4 characters per second, or 0.4 characters per frame. We doubled this limit to accommodate for the fast typing participants as well.

After assigning reading, writing or scrolling flags, we computed the proportionality vector from the counts of action flag from 5 seconds (50 frames). Finally, we used a winner-take-all strategy to assigned each 5 seconds episode a reading, writing, or scrolling label.

Gaze: Entropy-stability episodes
To define the entropy and stability episodes, we first overlaid a 50-pixels-by-50-pixels grid (hereafter referred to as grid) on the screen and we divided the whole debugging session into 10 second time windows (hereafter referred to as window). We then computed the proportion of time spent on each block in the grid. Using a two-dimensional proportionality vector, we computed the entropy and stability measures as follows:

**Entropy**: the entropy is calculated as the Shannon entropy of the proportionality vector. This measure tells us about the focus size of the participant. A value of 0 will show that the participant was looking at only one block on the grid. In other words, entropy measures the level of uncertainty of a random variable, which is, in our case, the objects looked at by the subjects. Theoretically, the highest possible entropy value is \( \log(\text{number of blocks in the grid}) \). In our case it is 2.76. This value will depict a uniform distribution of time over the grid. Thus, a high value of entropy would mean that the participant was looking at a wider range of objects on the screen; in other words, the participant had a higher size of focus (i.e. *not focused gaze*). Low entropy indicates that they mostly looked at few objects (i.e. *focused gaze*). Finally, it is important to point out that the “focus size”, in theory, is not at all related to “attention level”. It merely captures the number of objects the participant is looking at in a fixed time window.

**Stability**: the stability is computed over two successive windows. This measure tells us about the similarity between the objects looked by the subject and the duration of the gaze. We compute the cosine similarity between the two windows. This value is bounded between 0 and 1 (both included). A stability value of 1 will depict that the participant was looking at the same set of objects during two consecutive time windows, while a stability value of 0 will show that the participant was looking at completely different set of objects during two consecutive time windows.

Once, we had computed the entropy and stability for the participant, we applied a median cut to define focused and unfocused windows (based on entropy) and stable and unstable windows (based on stability). Finally we used a run-length encoding to define four types of gaze episodes: **focused-stable, focused-unstable, unfocused-stable, and unfocused-unstable**.

Gaze Transitions
We divided the whole Eclipse IDE into 8 functional Areas Of Interest (AOI) as follows.

1. **Variable View (VV)**: During debugging, allows changing the value of a variable to test how your program handles a particular value or to speed through a loop.
2. **Debug View (DV)**: Manages the debugging or running of a program in the workbench.
3. **Project Explorer (PE)**: Provides a hierarchical view of the artifacts in the workbench.
4. **JUnit (JU)**: Lists the unit tests to be passed by the main Java class.
5. **Exercise View (EV)**: Allows seeing the coding, saving, testing, and progress.
6. **Problem (Pr)**: Shows the errors and warnings raised by the Java Compiler.
7. **Console (Cn)**: Shows the output of the code.
8. **Code (Co)**: Panel where the code is written.

We then computed the gaze shifts from one AOI to another and grouped them following pre-defined categories: Locate problem, Locate variable, Hypothesis verification, Fix problem, Read code.

**Results**

Performance, unit-tasks, entropy-stability episodes
1) While in a main method running episode the success is positively correlated to the time that the participants spent as focused but unstable. 2) While in a junit test running episode the success is positively correlated to the time that the participants spent as focused and stable. 3) While in a junit test running episode the success is negatively correlated to the time that the participants spent as unfocused and unstable. Table 1 shows the
different linear models, and the Figure 1 shows the significant correlations as mentioned at the starting of this paragraph.

Table 1: Linear models for the success using different unit-tasks and entropy-stability episodes. Var. = variable; est. = estimate; JU = Junit; Fo = focused; Uf = unfocused; S = stable; Us = unstable; * = p-value (0.05); ** = p-value (0.01); *** = p-value (0.001)

<table>
<thead>
<tr>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val</th>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val</th>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.</td>
<td>3.8</td>
<td>0.9</td>
<td>4.2***</td>
<td>Int.</td>
<td>2.9</td>
<td>0.9</td>
<td>3.1***</td>
<td>Int.</td>
<td>2.6</td>
<td>1.0</td>
<td>2.5**</td>
</tr>
<tr>
<td>JU</td>
<td>0.7</td>
<td>1.2</td>
<td>0.5</td>
<td>JU</td>
<td>-2.3</td>
<td>1.2</td>
<td>-1.9</td>
<td>JU</td>
<td>2.7</td>
<td>1.4</td>
<td>1.9</td>
</tr>
<tr>
<td>Main</td>
<td>-1.5</td>
<td>1.1</td>
<td>0.5</td>
<td>Main</td>
<td>0.5</td>
<td>1.1</td>
<td>0.5</td>
<td>Main</td>
<td>0.7</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>FoUs</td>
<td>-6.1</td>
<td>8.6</td>
<td>-0.7</td>
<td>FS</td>
<td>16.3</td>
<td>8.1</td>
<td>2.1*</td>
<td>UfS</td>
<td>6.5</td>
<td>7.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Main:FoUs</td>
<td>16.5</td>
<td>8.0</td>
<td>2.1*</td>
<td>Main:FS</td>
<td>-0.9</td>
<td>8.9</td>
<td>-0.1</td>
<td>Main:UfS</td>
<td>-6.9</td>
<td>10.6</td>
<td>-0.6</td>
</tr>
</tbody>
</table>

Figure 1: Debugging success for the different gaze episodes and unit-tasks, the line shows the linear model.

Performance, unit-tasks, transitions
1) While in a junit test running episode the success is positively correlated to the proportion of locate problem transitions. 2) While in a junit test and main method running episode the success is positively correlated to the proportion of locate variable transitions. 3) While running the main method the main sub-task was to produce a desired output this could lead to "hypothesis generation - verification" loop and they go back and forth between code and console many times. 4) While in a junit test running episode the success is correlated to the proportion of fix problem transitions. Table 2 shows the different linear models, and the Figure 2 shows the significant correlations as mentioned at the starting of this paragraph.

Table 2: Linear models for the success using different actions and transitions. Var. = variable; est. = estimate; err. = error; JU = JUnit test; LP = locate problem; LV = locate variable; HV = hypothesis verification; FP = fix problem; * = p-value (0.05); ** = p-value (0.01).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.</td>
<td>5.3</td>
<td>0.9</td>
<td>5.4**</td>
<td>Int.</td>
<td>3.6</td>
<td>0.91</td>
<td>3.9**</td>
<td>Main</td>
<td>-1.1</td>
<td>1.3</td>
<td>-0.8</td>
</tr>
<tr>
<td>JU</td>
<td>-2.3</td>
<td>1.3</td>
<td>-1.1</td>
<td>JU</td>
<td>-2.2</td>
<td>1.1</td>
<td>-1.9</td>
<td>LV</td>
<td>8.9</td>
<td>7.6</td>
<td>1.1</td>
</tr>
<tr>
<td>LP</td>
<td>-7.2</td>
<td>8.6</td>
<td>-0.8</td>
<td>Main:LV</td>
<td>18.7</td>
<td>9.2</td>
<td>2.1*</td>
<td>Main:LP</td>
<td>22.2</td>
<td>11.0</td>
<td>2.1*</td>
</tr>
<tr>
<td>Main:LP</td>
<td>12.3</td>
<td>12.5</td>
<td>0.9</td>
<td>Main:LV</td>
<td>21.1</td>
<td>9.5</td>
<td>2.2*</td>
<td>Var.</td>
<td>4.5</td>
<td>0.8</td>
<td>5.4**</td>
</tr>
<tr>
<td>Int.</td>
<td>4.7</td>
<td>1.1</td>
<td>4.3**</td>
<td>Main</td>
<td>-2.7</td>
<td>1.1</td>
<td>-1.0</td>
<td>Main</td>
<td>-1.0</td>
<td>1.3</td>
<td>-0.7</td>
</tr>
<tr>
<td>JU</td>
<td>-0.3</td>
<td>1.3</td>
<td>-0.2</td>
<td>JU</td>
<td>-3.1</td>
<td>1.4</td>
<td>-1.8</td>
<td>FP</td>
<td>-1.0</td>
<td>8.4</td>
<td>-0.1</td>
</tr>
<tr>
<td>HV</td>
<td>0.8</td>
<td>7.0</td>
<td>0.1</td>
<td>FP</td>
<td>-1.0</td>
<td>8.4</td>
<td>-0.1</td>
<td>Main:FP</td>
<td>20.7</td>
<td>8.4</td>
<td>2.4**</td>
</tr>
<tr>
<td>Main:FP</td>
<td>11.9</td>
<td>11.3</td>
<td>1.0</td>
<td>JU:FP</td>
<td>36.4</td>
<td>12.8</td>
<td>2.8**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICLS 2018 Proceedings © ISLS
Performance, actions, entropy-stability episodes

1) While in a writing episode, the successful participants are focused stable. 2) While in a scrolling episode, the successful participants are unfocused unstable. 3) While in a reading episode, the successful participants are focused unstable. Table 3 shows the different linear models, and the Figure 3 shows the significant correlations as mentioned at the starting of this paragraph.

Table 3: Linear models for the success using different actions and entropy-stability episodes. Var. = variable; est. = estimate; Fo = focused; Uf = unfocused; S = stable; Us = unstable; * = p-value (0.05); ** = p-value (0.01)

<table>
<thead>
<tr>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val.</th>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val.</th>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.5</td>
<td>8.1</td>
<td>4.4**</td>
<td>Int</td>
<td>5.4</td>
<td>0.9</td>
<td>5.8**</td>
<td>Int</td>
<td>4.3</td>
<td>1.1</td>
<td>3.8**</td>
</tr>
<tr>
<td>read</td>
<td>1.4</td>
<td>1.4</td>
<td>1.0</td>
<td>read</td>
<td>-0.7</td>
<td>1.4</td>
<td>-0.4</td>
<td>read</td>
<td>-2.8</td>
<td>1.4</td>
<td>-1.8</td>
</tr>
<tr>
<td>scroll</td>
<td>1.3</td>
<td>1.1</td>
<td>0.1</td>
<td>scroll</td>
<td>-2.3</td>
<td>1.2</td>
<td>-1.8</td>
<td>scroll</td>
<td>1.9</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>write</td>
<td>-1.9</td>
<td>1.1</td>
<td>-1.5</td>
<td>write</td>
<td>-0.3</td>
<td>1.3</td>
<td>-0.2</td>
<td>write</td>
<td>0.2</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>FS</td>
<td>9.8</td>
<td>6.3</td>
<td>1.5</td>
<td>UfUs</td>
<td>-6.8</td>
<td>6.1</td>
<td>-1.1</td>
<td>UfS</td>
<td>2.1</td>
<td>7.5</td>
<td>0.2</td>
</tr>
<tr>
<td>read:FS</td>
<td>-1.9</td>
<td>9.5</td>
<td>-1.3</td>
<td>read:UfUs</td>
<td>5.6</td>
<td>9.7</td>
<td>0.5</td>
<td>read:UfS</td>
<td>21.6</td>
<td>9.8</td>
<td>2.2*</td>
</tr>
<tr>
<td>scroll:FS</td>
<td>-4.3</td>
<td>8.9</td>
<td>0.1</td>
<td>scroll:UfUs</td>
<td>18.3</td>
<td>7.9</td>
<td>2.3*</td>
<td>scroll:UfS</td>
<td>-0.1</td>
<td>10.6</td>
<td>0.1</td>
</tr>
<tr>
<td>write:FS</td>
<td>15.4</td>
<td>7.4</td>
<td>2.1*</td>
<td>write:UfUs</td>
<td>2.6</td>
<td>9.0</td>
<td>0.2</td>
<td>write:UfS</td>
<td>-1.9</td>
<td>10.2</td>
<td>-0.9</td>
</tr>
</tbody>
</table>

Performance, actions, transitions

During reading, writing, and scrolling episodes success is positively correlated to locate problem, hypothesis verification, and locate variable transitions, respectively. Table 4 shows the respective linear models.

Table 4: Linear models for the success using different actions and transitions. Var. = variable; est. = estimate; LP = locate problem; LV = locate variable; HV = hypothesis verification; ** = p-value (0.01)

<table>
<thead>
<tr>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val.</th>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val.</th>
<th>Var.</th>
<th>est.</th>
<th>error</th>
<th>t-val.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.3</td>
<td>1.1</td>
<td>5.0**</td>
<td>Int</td>
<td>5.4</td>
<td>0.8</td>
<td>6.1**</td>
<td>Int</td>
<td>5.4</td>
<td>0.9</td>
<td>6.0**</td>
</tr>
<tr>
<td>read</td>
<td>-3.8</td>
<td>1.3</td>
<td>-1.8</td>
<td>read</td>
<td>-0.9</td>
<td>1.2</td>
<td>-0.7</td>
<td>read</td>
<td>0.3</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>scroll</td>
<td>0.009</td>
<td>1.4</td>
<td>0.1</td>
<td>scroll</td>
<td>-4.2</td>
<td>1.2</td>
<td>-1.4</td>
<td>scroll</td>
<td>0.009</td>
<td>1.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Discussion and Conclusions

In this contribution, we study the relation between the gaze patterns and the actions, each at two different levels, to explain the difference between successful and unsuccessful debuggers. We defined two gaze variables: focus-stability episodes and transitions; and two action variables: unit-tasks and reading/writing/scrolling episodes. Our results show that these variables, when combined together, can enable us to understand debugging approaches, that extend beyond the state-of-the-art “finding bugs” approaches. This is one way that can help programming educators to capitalize on the debugging tools to develop novel instruction to teach students how to debug (McCauley et al., 2008).

Combining the unit tasks (main method execution and JUnit test) with the gaze variables, we observed that the students follow two different trajectories during the two different unit tasks. While in the main method execution episode the successful students locate the problem in the code and perform a hypothesis generation-verification cycle. This results in a focused but unstable gaze for two reasons. First, they have to look at different small parts of the program to locate the problem; and second, to perform the hypothesis generation-verification they have to repeatedly switch between the code and the output resulting in looking at two different (unstable gaze) but small sets of objects looked at (focused gaze). On the other hand, the JUnit test descriptions are designed to be precise. The successful students use this description to locate the problem causing variable and to fix it. To do so, the students often concentrate on a small part of program (focused gaze) for a longer duration (stable gaze).

Combining the actions (reading, writing, scrolling) with the gaze variables, the data revealed three different strategies from the successful students. First, while reading, the successful students try to locate the problem as a result, and as explained before, they have a focused-unstable gaze. Second, while writing, the successful students fix the problem using a hypothesis generation-verification cycle, and in turn they exhibit a focused-stable gaze, similar to the previous paragraph. Finally, when scrolling the students perform a typical search for the problematic variable and hence have a unfocused-unstable gaze. One can argue that this relation between gaze and action patterns is not surprising. However, with our analysis, we have shown an automatic way of detecting the successful action-gaze pattern. In future, this can be leveraged to design and develop real-time automated tools to help the students while debugging a program.

The findings reported in the second paragraph are not obvious and were reached by analyzing gaze data. However, many researchers might argue that the findings reported in the third paragraph were obvious, but we would argue that those findings are intuitive but not obvious. As it can be seen through the study design, we let the students to write the code, which makes eye-tracking analysis not-straightforward, since the stimulus changes with time and is different for each student.

In a nutshell, we showed that combining gaze data with actions (reading, writing, scrolling) and unit tasks (main method execution and JUnit test) help us to further understand the cognition that underlies debugging a program and discover debugging patterns of successful students. Moreover, allowing students to “find and fix the bug” helped us to capture more in-depth data in terms of actions and gaze during the debugging task. Adapting to the technological advancements and using eye-tracking to study successful and
unsuccessful programmers provide important insights for improving teaching and learning how to teach debugging as a skill successful programmers need to have it. The results are interesting enough to pursue further research to observe how and when novices build basic knowledge about variables and expressions, various mental models, as well as what misconceptions cause what type of programming errors.

References


Teacher Practice Spaces: Examples and Design Considerations

Justin Reich, YJ Kim, Kevin Robinson, Dan Roy, and Meredith Thompson
jreich@mit.edu, yjk7@mit.edu, krob@mit.edu, danroy@mit.edu, meredith@mit.edu
Massachusetts Institute of Technology

Abstract: Teacher practice spaces are learning environments, inspired by games and simulations, that allow teachers to rehearse for and reflect upon important decisions in teaching. Practice-based teacher educators use a variety of approaches to simulation in methods courses and other professional learning opportunities, and existing simulations often attempt to holistically replicate authentic teaching conditions. We extend this work by developing new kinds of practice spaces that do not attempt to fully simulate teaching, but rather offer playful and creative opportunities for novice teachers to develop skills and dispositions valuable for teachers. We summarize six different practice spaces developed through design research, and then articulate a set of design considerations emerging from this work to expand the genre of pedagogies of enactment in teacher professional development.

Introduction
Every great teacher knows that skill development requires practice (Ball & Forzani, 2009); ironically, teachers themselves have limited opportunities to practice important teaching moves in low-stakes settings. In a comparative study of teachers, social workers and therapists, Grossman and colleagues (2009) conclude that “prospective teachers have fewer opportunities to engage in approximations that focus on contingent, interactive practice than do novices in the other two [helping] professions.” Currently, teacher candidates primarily learn in two spaces: Socratic seminar rooms in education schools (or lecture-heavy workshops for in-service professional development) and practicum classrooms. The former affords discussion and the latter affords immersion into the challenges of teaching, but a third space—a practice space—is needed that combines the authenticity of the practicum classroom with the scaffolding of the graduate school seminar room.

In this article, we summarize a set of design experiments creating teacher practice spaces—learning environments inspired by games and simulations where teachers can rehearse for and reflect on important decisions in teaching. These practice spaces build on rich literature and examples of simulation in teacher education. Most of the existing efforts at simulation in teacher professional learning aim for comprehensiveness; they attempt to immerse students in experiences that closely approximate the full complexity of teaching experiences. For instance, the Sposato Graduate School of Education at the Match Charter School has done inventive work with mock teaching among colleagues (Match 2017), and Mursion creates mixed reality simulations where teacher-learners teach mock lessons with virtual student avatars controlled by a remote operator (Mursion, 2017). These examples aim to holistically simulate the complexity of teaching. By contrast, many of the practice spaces in our design experiments do not attempt to recreate the full complexity of teaching. Instead, these projects explore playful mechanics that allow teachers to approximate skills and dispositions that are valuable for teaching without trying to simulate the whole. To borrow an analogy from sports, much of the work in practice-based teacher education has focused on creating scrimmages—experiences that closely model the complete endeavor. In our practice spaces, we have focused on creating drills: activities that facilitate targeted development of useful skills for an activity without trying to recreate the entire activity. Our research into practice spaces explores playful mechanics that allow teachers to learn skills valuable for teaching through pedagogies of enactment.

In what follows, we describe several important lines of research that inform our design efforts: work in practice-based teacher education, pedagogies of enactment, and clinical simulation in teaching. We then summarize research from six of our existing practice spaces in various stages of development, ranging from Baldermath (a bluff the judge game about fractions), to Committee of N (a tarot deck for school design), to TeacherMoments (a handheld teaching simulation app). Finally, we articulate a set of design consideration that emerge from examining the full set of these practice spaces, and offer suggestions for further exploring this promising domain.

Background and context
A movement towards practice-based teacher education has grown over the last two decades and expanded greatly in the last ten years. The core tenant of this movement is shifting teacher education away from an emphasis on specifying necessary knowledge for teaching towards specifying teaching practices that include
both knowing and doing (McDonald, Kazemi, Kavanaug, 2013). The word practice takes on multiple meanings in teacher education—practices are the set of activities that teachers employ in their work and practice is the rehearsal and preparation required to expertly deploy those activities in classrooms (Lampert, 2010). Practice-based teacher education emphasizes helping novice teachers learn specific instructional activities (practices) in large part through rehearsing (practicing) them outside the classroom before employing them in the classroom. This approach is often contrasted with teacher-learning experiences in education school classes or school-based professional development, where teacher-learners discuss theory or instructional approaches (or listen to lectures about these topics), but have few opportunities to enact them.

The proper grain size of these practices is debated. Lemov (2010) provides instruction and video coaching on very specific classroom teaching moves, mostly designed to maintain the continuous attention and time-on-task of students. Lampert and Graziani (2009) argue for a broader frame, where instructional activities are as defined activities with a beginning, middle, and end but variable in how they may unfold in differing contexts. Lemov’s teaching moves take seconds, Lampert and Graziani’s instructional activities unfold over minutes. Also debated is the degree to which these instructional activities should be understood as standardized versus situated: to what extent are their universally better or worse teaching strategies and to what extent does the effectiveness of any given teaching decision depend on contextual factors of setting and students? In frameworks that view good teaching as universal, practice of teaching moves looks like an actor’s rehearsal of a scripted scene, where faithful enactment of the defined elements of the scene is the goal (e.g. Match 2017) . In frameworks that view good teaching as situated, practice looks more like rehearsal for improvisational theatre, where key tenets of good performance are interpreted differently for different situations (e.g. Self 2015). In our practice spaces, we are biased towards designs that have the potential to accommodate both perspectives.

Teacher educators have developed a variety of approaches for rehearsing instructional practices. Grossman and colleagues (2009) define a three-part framework for learning and apprenticeship in the helping professions (teaching, social work, etc.): representation, decomposition, and approximation. Teacher educators share representations of teaching (their own modeling, examples of lesson plans, videos of teaching), then decompose and highlight important elements (pacing, classroom management strategies, questioning techniques) from those representations. Teacher-learners then approximate these practices by writing up practice lesson plans or teaching mock lessons. The development of strategies for incorporating these approximations in teacher education are sometimes called pedagogies of enactment (Kazemi, Franke, & Lampert, 2009).

The last decade has seen tremendously productive design research into these pedagogies of enactment. The Core Practices Consortium is a group of teacher educators from multiple disciplines exploring important practices in teaching and new strategies for rehearsing those practices (McDonald, Kazemi & Kavanaug, 2013). Kazemi, Franke, and Lampert (2009) define a cycle of enactment where novices participate in lessons with particular instructional activities as students, teacher-learners then analyze this representation and identify strengths and alternative pathways, novices prepare a plan for teaching with the protocol, and then teacher-learners take turns rehearsing their lesson with colleagues acting as students. The Sposato School of Education promotes similar cycles of scripting instructional moves, practicing with colleagues, and then receiving coaching from peers and mentors. Field-mediated experiences are an innovative approach to practice in methods courses, where entire methods classes visit the classrooms of cooperating teachers for immersive coaching experiences with small groups of K-12 students (CITE XXX). Mursion is a mixed-reality medium where teachers practice teaching skills in front a computer monitor with a series of student avatars (Mursion, 2017). The student avatars are controlled as “digital puppets” by a remote actor who can see and hear the teacher through networked video cameras. Dotger (2013) designs clinical simulations, adapted from medial education, for pre-service students. Teacher-learners interact with actors who are trained with the background story and motivations of a parent, student or colleague with some particular problem or dilemma. Through preparation, roleplay, individual reflection, and peer discussion, novice teachers rehearse for challenging situations in low-stakes settings. Self (2016) has adapted these clinical simulations for addressing issues of culturally responsive teaching. From specific teaching moves, to complex instructional activities, to conversations with parents and students, research into pedagogies of enactment shows promising results in helping novice teachers be better prepared with specific techniques and strategies as they begin classroom teaching.

How widely these approaches are adopted is an open question. No comprehensive study measures the adoption of these new practice-based approaches across GSEs or school districts. As perhaps the newest discipline-based field in teacher education, pedagogies of enactment are well-represented in Computer Science teacher education. In the Exploring Computer Science curriculum (Margolis, Google & Chapman 2015), novice teachers regularly engage in computer science learning activities, create lessons, and teach them with colleagues. In their report on the challenges of teacher education, (TNTP 2015) surveyed teachers in three large districts and found that 17, 27 and 38 percent of teachers, respectively, reported practicing teaching as part of
professional development “sometimes” or “often”. In a charter management organization they surveyed, 82% of teachers reported practice-based professional development happened “sometimes” or “often.” Our sense is that adoption of these practices is growing, but still a limited part of most typical teacher-education and in-service professional development programs.

Our practice spaces are inspired by all of this important work into pedagogies of enactment, and our research attempts to expand new genres of approximation and open new directions in pedagogies of enactment. We observe that most efforts at practice in teacher education aim to approximate as completely as possible the experience of teaching a classroom. In many of our practice spaces, we deliberately relax the constraint of attempting to comprehensively recreate teaching in order to create a wider design space for practice activities. Some of our practice spaces are slices of teaching where we abstract away much of the complexity of the classroom in order to focus on some specific dimension, such as our interactive video case studies where teachers encounter short vignettes of difficult teaching circumstances. Others of our practice spaces engage teacher-learners in non-teaching activities that help them develop skills and dispositions that are useful for teaching. Analogies to sports and musical instruction can be helpful. When training young violinists, music teachers often use bow games: silly songs where violin-learners sing and wave their bow with specific motions while maintaining the correct grip on the bow handle. Young soccer athletes play games such as keep-away to develop ball-handling skills. A violinist will never waive her bow maniacally above her head in a recital, and a soccer player will never play keep-away during a match, but these drills isolate particular skills for development that are then re-integrated—ideally with greater competency—into the complex assemblage of the whole activity. Our teacher practice spaces have the same goal, to introduce new kinds of drills into teacher education, and if these drills prove successful, then they could be added alongside discussions of theory, holistic simulation, and field placements in the repertoire of teacher educators.

Designing teacher practice spaces
Over the past two years, we have used design-based research methods (Easterday, Rees-Lewis, & Gerber, 2014) to develop a diverse set of practice spaces. In our design process, we typically begin with construct development, where we identify a skill that we want novice teachers to develop and then consider what they skill looks like when expertly deployed in the classroom. We then prototype playful experiences that allow novice teachers to enact these skills or practices. The ideal design team includes a combination of teachers and teacher educators with a deep understanding of the targeted constructs along with game designers who bring encyclopedic knowledge of existing game mechanics. We iteratively improve new practice spaces through frequent playtests among our lab members along with regular lab-based playtests with pre-service teachers, in-service teachers, and teacher educators. We refine our practice spaces and develop curriculum with which to embed them through field tests in teacher preparation programs or in-service teacher professional development. In the projects discussed below, our field test partners include MIT’s Scheller Teacher Education Program, West Virginia University, the College of St. Scholastica, Code.org, Exploring Computer Science, Mobile CSP, and the Hartford Magnet School in Connecticut, all of whom provided invaluable feedback on the projects.

Ideally, the design process of each of our practice space projects would lead to development of four components. First, we develop a practice space with a play mechanic encoded in rules, playing cards, or software that are made available through open source software or openly licensed materials (tsl.mit.edu/practice). Second, we develop authoring tools for our practice spaces so that teacher educators can modify our materials, change rules, and create their own scenarios. For a card game, we might make PDF files of cards available so that teacher educators can edit them. For our simulation platform for handheld devices, we aspire to create software authoring tools for teacher educators to create their own scenarios. Third, we create curriculum examples for each practice space; none of our practice spaces are intended to be entirely self-contained learning experiences. At a minimum, every practice space experience requires some kind of debriefing or reflection experience to be meaningful, and some of our practice spaces benefit from advanced preparation or interleaved periods of play, reflection, learning, and more play. Fourth, we publish scholarly articles that describe the design principles behind our experiences and provide evidence of their efficacy. Below, we summarize the development and research of six very different teacher practice spaces before describing principles that emerge from examining the full set.

Six example practice spaces
Below we detail six practice spaces. Playable demos, game materials, curriculum suggestions and other resources can be found at tsl.mit.edu/practice.
Baldermath
Baldermath is a bluff-the-judge game about teaching fractions (Pershan, Kim, Thompson, & Reich 2017), co-designed by the author of the MathMistakes.org blog (Pershan, 2017), an online space where teachers discuss interesting errors from math students. To play the game, a judge leaves the room, and four players are given a homework problem taken from a fourth-grade classroom. One contestant is given an actual piece of student work for the problem, completed by a student with an incorrect or incomplete understanding of the problem. This contestant copies the work into her own hand, and then invents a rationale for why the student thought she was correct. The other contestants invent incomplete or incorrect answers to the problem as well as rationales. The judge returns to the room, the contestants roleplay as students and explain their concocted rationales along with details of their (fabricated or real) student work. The judge then guesses which is the “real” student work. As with Balderdash or the Wait, Wait Don’t Tell Me News Quiz, guessing correctly is fun for the judge and guessing wrong is fun for the winning contestant.

Math education research shows that novice teachers typically engage in three unhelpful practices when looking at student work: fixating on whether the answer is correct, making assumptions about the demographics or intelligence of the student, and making assumptions about the quality of instruction (Pershan, Kim, Thompson, & Reich 2017). By contrast, the most useful practices are looking closely at specific details of student work and making inferences and hypotheses about student thinking and understanding. In Baldermath, the game mechanic naturally guides teachers towards engaging these productive practices and eschewing the unhelpful ones; the game is only fun and winnable by thinking carefully about student thinking and representations of that thinking. In debriefing the game experience, facilitators can show teachers how the practices developed playfully during Baldermath can be productively applied to looking at student work in homework and classwork settings.

Metarubric
Metarubric is a playful examination of the challenges with using rubrics to evaluate complex performance (Kim, Rosenheck & Reich, In Submission). Participants select a movie by consensus (such as Titanic) and then briefly create movie posters for the movies. Participants then create rubrics for the posters, and take turns using their different rubrics to grade the posters. In a follow-up round, players develop a rubric for the rubrics—the metarubric—and then take turns grading the rubrics themselves. In conversations between rounds, players typically observe that their favorite posters do not necessarily get the highest rubric scores, and that most rubrics undervalue a component of their poster that they felt was important.

In most teachers lived experience, rubric usage unfolds over weeks: projects are developed, rubrics are designed, students do the projects, and teachers grade them with rubrics. In Metarubric, participants go through this cycle several times within an hour. By collapsing the time of this cycle, participants viscerally encounter tensions with the problematic elements of rubrics. By empathizing as learners experiencing how rubric scores imperfectly map onto the worthy qualities of performance assessments, teachers expand their thinking about how to better align the goals of a learning experience with the assessment criteria for performance assessment.

TeacherMoments
TeacherMoments is a simulation designed for handheld devices, where participants are immersed in short vignettes of teaching life rendered in text, animation or video, and participants respond to “triggers” with text or improvisational audio responses (Owluakporie, Thompson, Robinson, & Reich, In Submission). In live-actor clinical simulations used in teacher education (Dotger, 2013), actors are trained to portray parents or students in a specific situation. Briefing books given to actors include the background of the character and situation, as well as a series of “verbal triggers” that actors are supposed to include in the conversation (such as “You only called me out because you are racist” or “But what will do you when my (autistic) son hugs someone at an inappropriate time?”). Since these actors are meant to create standardized situations, TeacherMoments tests the viability of encoding these interactions entirely in text and video. For instance, Dotger (2013) has developed a series of parent simulations, including one parent upset because a class is too hard; in TeacherMoments, we record six video sequences of an actor playing this parent. Novice teachers participating in the simulation are required to provide improvised audio responses after each recorded conversational turn. In Dotger’s live-actor role plays, his four goals for participants are that 1) they experience the interaction as authentic, 2) that the scenario generates a feeling of cognitive disequilibrium, 3) that participants demonstrate an ability to remain calm under pressure, and 4) that they can articulate some element of their teaching philosophy in response to the verbal triggers from actors. Our playtests suggest that these four goals are met within the experience of TeacherMoments, even though our “actor” is pre-recorded rather than live. Given that teachers...
may never meet a parent during their practicum experience, this application of TeacherMoments gives teacher-learners a chance to practice an important dimension of teaching before their induction period.

TeacherMoments is designed so that teacher educators can create different kinds of scenarios and case studies that allows teachers to rehearse different competencies. We have created a series of scenarios that help students address equity teaching practices in Computer Science instruction (Robinson & Reich, Forthcoming). Integrating culturally responsive pedagogies in computer science classrooms requires attention to curriculum design and classroom practices and routines, but the “last mile” of equitable teaching are the in-the-moment decisions made by teachers. Equity teaching practices that can shape in-the-moment teaching decisions include addressing preparatory privilege (the advantages of experience that many students, especially white and Asian boys, bring to CS classrooms), acknowledging students’ intersectional identities, and adopting asset framings. Pedagogies of enactment, such as TeacherMoments simulations, are particularly effective for surfacing what Dotger and Ashby (2010) calls “conditional inclusive ideologies.” These are beliefs about equitable teaching that novice teachers espouse in discussions but often fail to act upon in specific circumstances. For instance, in one of our equity scenarios, two students are assigned a paired activity and one student asks to be allowed to move on to more advanced work rather than work with his disengaged partner. Even teachers who articulate concerns about preparatory privilege in general, may choose to be more concerned with providing additional challenging work for the advanced students rather than addressing how his behavior may be marginalizing his partner. We find, following Dotger and Ashby, that pedagogies of enactment and participating in our interactive case studies are effective mechanisms for surfacing teacher’s conditionally inclusive ideologies and opening up new conversations about the competing values that shape teacher decision making.

Eliciting Learner Knowledge (ELK)

Eliciting Learner Knowledge (ELK) is a two-person online game, with one person role playing a teacher and another role playing a student Thompson, Roy, Wong, Reich, & Klopfert, Forthcoming). In the ELK platform, players have a conversation through a text-based, chat-like interface. This format has two potential advantages over in-person role plays: the transcript of the conversation allows for immediate reflection on specific, documented details in the exchange, and the conversation can occur when two people are not in the same location, as will be increasingly common as more teacher education happens online. Each round of the game focuses on a conceptual topic in science such as chemical reactions, evolution, or energy, or a topic in mathematics such as rational numbers, fractions, and proportions.

At the beginning of the game, each player receives instructions and a brief overview of the game; the person role-playing the teacher receives a learning objective and the person role-playing the student receives a learner profile with details of the concepts and misconceptions held by the the student being role-played. Players review the profiles, engage in a synchronous 7-minute conversation, and then both players take the same true/ false quiz as if they were the “student”. To encourage collaboration and communication between the players, the quiz is scored on 1) how well the student portrays the student profile, and 2) how well the teacher estimates the student’s understanding. ELK has two goals: to help preservice and in-service teachers understand questioning strategies and to learn about possible student misconceptions.

Committee of N

Committee of N is a design-based card game for exploring education history and policy through school design (Haas, Reich, Feely, Klopfert, 2016). Participants play as consultants charged with designing elements, such as a classroom design or graduation requirements, of a new high school. Each Committee of N deck includes eight of these design elements, along with different sets of “value cards” representing belief commitments from the fictional new school. Participants work in pairs, and each round they are dealt hands that include one school design element, and then three school values. Values can include purposes of education (e.g. assimilating immigrants or career/college readiness), theories of learning (e.g. behaviorism to constructionism), instructional methods (e.g. apprenticeship or flipped classroom). A pair might be asked to design the bell schedule for a school inspired by behaviorism, committed to vocational education, and emarored of project-based learning. Pairs create four to eight of these design elements, and then join up with several other pairs to create a school out of their elements. Teams then pitch these joint schools to a panel of “school committee” judges.

For many novice teachers heading out into the field for observations, many elements of school seem fixed and immutable. Committee of N helps novice teachers see that every practice, every fixture, every routine within in a school was designed at some point in history by people who held a set of values, and if we no longer hold those values we can design new school elements with new values. Not every change is equally easy— extra-curricular activities can be redesigned easily whereas most communities can only rotate through new school building after several decades--but recognizing that school elements were designed empowers novice
teachers to imagine how they might be designed anew. Many students adopt the heuristic of describing the “value cards” underlying the practices and fixtures they see in their school observations. The game mechanic underneath Committee of N is essentially the same mechanic as the Tarot: players create stories about the future guided by a series of arbitrary constraints, and by imaging different possible futures, players can reflect on which futures they would like to try to bring about in the world.

Motivation station
Motivation Station is an in-person card game that creates scenarios for novice and experienced teachers to practice applying principles of cognitive science to motivating students. The gameplay mechanic is similar to Apples to Apples. Each round one participant acts as a judge, and draws cards for a particular student and a particular scenario involving a dilemma of student motivation. Other players hold hands of cards with cognitive science-based principles for motivating students, and each round they select which principle they think would best address the scenario. If the judge selects a player’s motivation principle, that player then needs to act out how she would use that principle to address the situation. The judges evaluate the effectiveness of the response, creating natural opportunities for discussion and comparison.

Design considerations in teacher practice spaces

Authenticity versus playfulness
One useful gradient for analyzing our six practice spaces is the degree to which they approximate authentic teaching practices. In our set of six practice spaces, TeacherMoments is closest to the authentic approximation—it is the least game-like and the most simulation-like. ELK might be quite authentic for a fully online teacher, but for most classroom teachers the scenarios feel authentic but the interface is an artifice. By contrast, a practicing teacher will never need to fabricate incorrect student work as in Baldermath or use a rubric to evaluate rubrics as in MetaRubric. We find in our feedback across playtests of these different environments that typically the closer an activity replicates authentic teaching practice, the less likely it is to feel playful and fun. It may still feel authentic, challenging, and worthwhile, but novice teachers typically do not experience simulation as playful. Releasing the constraint of authenticity opens up an interesting design space, of playful activities where novice teachers do work that is relevant to teaching while not replicating the exact tasks of teaching. The literature of game-based learning gives us many reasons to be optimistic about the potential of包括 more playfulness in teacher education: the intrinsic motivation offered through play, the freedom to take risks under the guise of projective identities, the opportunity to experiment with new pedagogical commitments or tactics in low-stakes settings, and revealing of the operation of complex systems by modeling elements of those systems in games (Gee, 2007).

Surfacing problems versus scaffolding practice
Another gradient across our practice spaces is the degree to which participants develop new skills during gameplay. In MetaRubric, participants discover all kinds of tensions and dilemmas in the use of rubrics, but the game play does not necessarily scaffold improvements in practice. Our hypothesis is that teachers will improve their assessment practices because they empathize more closely with how students experience rubrics and recognize more deeply the tensions in designing effective rubrics. By contrast, the game development process for Baldermath unearthed a clearly defined construct for looking at student work—with well-defined productive and unproductive practices, and the mechanics of the game naturally guide participants away from unproductive practices and towards productive ones. A simple debrief at the end of the experience may be sufficient for novice teachers to consciously adopt these new practices. Committee of N doesn’t scaffold a teaching practice per se, but participants learn a useful heuristic—the idea of values underlying a school design element—that can help them better understand the constraints of their context. For the practice spaces that surface problems rather than scaffold practice, we hypothesize that other curriculum elements will need to provide scaffolding in developing novice teacher skills, and the games provide empathy and insight. Across our design efforts, we have found that rigorous construct definition—carefully defining the characteristics of effective practice—is essential to creating games or simulations that scaffold particular practices or approaches.

Considering reusable mechanics
Many of our practice spaces adopt existing game mechanics. TeacherMoments draws on the pedagogy of clinical simulations, Baldermath is a variation on Balderdash, Motivation Station is a variation on Apples to Apples, and Committee of N is a variation on the Tarot. A key element to both ELK and MetaRubric is the acceleration of time. In MetaRubric, players conduct a full cycle of rubric development and use over minutes
rather than days or week, in order to more viscerally confront the tensions of rubric use. We imagine that in future practice space design, these kinds of new and established mechanics will be re-used across different variations of practices spaces, and drawing on existing mechanics is a promising approach to rapidly prototyping new practice spaces and exploring the possibilities of the design space.

Role-playing students playfully builds empathy
Our participants generally enjoy the chance to role-play as students. Participants find it challenging and provocative to think like a confused student in Baldermath, or to reflect on how it feels to receive rubric scores on a complex performance in Metarubric, or to take on the role of a student with incomplete understanding in ELK. Role-playing as students allows novice teachers to reflect on the curse of expertise and dimensions of the novice-expert divide, and roleplay builds empathy with confused students, frustrated parents, and teacher colleagues doing their best in the complex circumstances in schools. Playfully exploring new identities, such as a new and vehement commitment to behaviorism in a Committee of N design round, allows teachers to try on new pedagogical commitments in low-stakes spaces. In our early design work, we thought of student roleplaying as a transactional part of game design: some players role-played as students to help others being teachers. We now think of student roleplaying as an element that is enjoyable and valuable in its own right.

Gathering evidence of efficacy
Lab-based playtests and small field tests of practice spaces are relatively simple to conduct. In most of our playtests, the majority of the data we collect are artifacts of game play along with participants subjective impression of their experience. These are sufficient to iteratively improve the practice spaces, especially in optimizing for playability and fun, but we have only limited evidence of how our practice spaces are actually changing teacher behavior. Gathering this kind of evidence will require deeper partnerships with more teacher education and professional development programs, and greater investment in collecting observational data about teacher practice before and after gameplay. We have conducted some experiments with simpler near-transfer tasks. For instance, in Baldermath playtests we have had participants look at student work using a Notice/Wonder protocol pre- and post-game play; we then look for evidence that participants notice more interesting and relevant details and make better hypotheses about student thinking after game play. These near-transfer tasks strike a balance between the purely subjective and impressionistic data collection from many of our playtests and the very high hurdles of using in-depth field research to gather evidence of behavioral change. As this line of research matures, we will need to better understand how well the subjective measures we collect during the game design process correlate with measures of actual behavior change—do the practice spaces that teachers tell us they enjoy and found informative or enlightening actually lead to changes in behavior?

Drills and scrimmages: Isolated skills and complex assemblages
Overall, our prototype practice spaces have received a positive response from the teachers and teacher educators who play them. They are playful, engaging, and provocative, and they draw teachers’ attention to the details of in-the-moment teaching decisions: as one participant said during a debriefing, “I used to think of good teaching in terms of good and bad lessons. I now think of good teaching in terms of good or bad 15 seconds.” We do not intend these practice spaces to replace more immersive approaches to teaching simulations: mock lessons with colleagues and clinical simulations are critical to the future of practice-based teacher education. That said, releasing ourselves from the constraint of authenticity has opened our design space for inventing new kinds of practice-spaces. As Grossman and colleagues (2009) explain, one of the tensions with pedagogies of approximation is how much to approximate. Teaching requires deploying skills simultaneously in a complex assemblage—in a real classroom a teacher is simultaneously watching the clock, evaluating student attentiveness, drawing on knowledge about student relationships and competencies, and making constant decisions about pacing, behavior management, and student agency. Each of these teaching decisions is intimately entangled with the others, so a tension emerges between isolating skills out of the complex assemblage for practice (since the isolated skill is easier to address than the whole assemblage) and recognizing that none of these elements are actually isolated in real classrooms. Our development of new teacher practice spaces has reinforced our confidence in the potential for isolating particular skills. Just as athletes conduct drills to isolate skills that can be integrated back into the complexity of a whole game, we hope that novice teachers will integrate the skills developed in practice spaces back into the complex whole of their teaching practice. To what extent that re-integration actually occurs and teachers actually change their behavior is the most important question for this line of research to next address.

References


Owho-Ovuakporie, K., Thompson, M., Robinson, K., & Reich, J. (In Submission). Teacher Moments: An online platform for preservice teachers to practice and reflect on parent-teacher conversations.


Thompson, M., Roy, D., Wong, T., Reich, J., Klopfer, E. (Forthcoming) Moving from diagnosis to understanding students ideas through the game of Eliciting Learning Knowledge (ELK). To be presented at the National Association for Research in Science Teaching, Atlanta, GA.


**Acknowledgments**

We gratefully acknowledge the support of the Woodrow Wilson National Fellowship Foundation and Google in the development of our practice space research.
Mapping Networks to Help Education Leaders Gain Insights Into Complex Educational Systems

Robbin Riedy, University of Colorado Boulder, robbin.riedy@colorado.edu
Katie Van Horne, University of Colorado Boulder, katie.vanhorne@colorado.edu
Philip Bell, University of Washington Seattle, pbell@uw.edu
William R. Penuel, University of Colorado, Boulder, william.penuel@colorado.edu
Tiffany Neill, Oklahoma Department of Education, tiffany.neill@sde.ok.gov
Sam Shaw, EdReports, sshaw@edreports.org

Abstract: Developing an understanding of education as a complex, adaptive system can be useful in helping leaders understand the impact of various interventions on the system as a whole. This paper describes the design of two learning opportunities, which draw on Actor Network Theory, for assisting state science coordinators in exploring complex educational systems in their state. We respond to a call for research on how representational tools can help education leaders develop understandings of complex systems. The authors conducted a design study to address the question, how do state science education leaders use representations of complex systems to understand their networks and identify leverage points for change?

Introduction and framework

Understanding large educational systems as complex adaptive systems can provide insights that can help leaders anticipate both intended and unintended consequences of policies they devise (Maroulis et al., 2010). Tools such as agent-based models and network maps have the potential to facilitate this understanding (Daly & Finnigan, 2010; Penuel, Harrison, et al., 2015). Thinking about complex systems requires both domain knowledge and an understanding of systems concepts, which may violate pre-formed intuitions. However, developing this understanding may prove difficult, as “learning about these systems challenge cognitive, metacognitive, and social resources” (Hmelo-Silver & Azevedo, 2006, p. 58). Designing for change in educational policy requires paying specific attention to the implementation of interventions (Penuel, Fishman, Cheng, & Sabeli, 2011), the relationships between and coordination among different actants in a system (Jackson & Cobb, 2013), and the coherence within the system (Forman, Stosich, & Bocala, 2017; Fuhrman, 1993; Linn, Kali, Davis, & Horwitz, 2008). Still, a complex systems approach can be helpful in identifying specific leverage points, or places where it may be especially productive to focus efforts, in systems, because of the disproportionate effect interventions at those points may have. Policy efforts that focus on these specific leverage points can be helpful in achieving large scale change (Bryk, Gomez, Grunow, & LeMahieu, 2015).

Due to their potential utility in helping education leaders re-imagine and re-organize educational systems, there is a need to further explore the pedagogy, and the design of activities intended to use representational tools to help teach complex systems thinking related to change in educational systems (Jacobson & Wilensky, 2006). Work by constructivist learning theorists has demonstrated the utility of simulations and visual representations for learning. For example, NetLogo, Logo and Turtle programming have been used to help learners foster a deeper understanding of networks (Wilensky & Rand, 2015; Colella, Klopfer & Resnick, 2001; Papert, 1972). In addition, research in the learning sciences has already identified several aspects of good pedagogy for teaching about complex systems, such as the importance of scaffolding, collaboration, discussion, and reflection in learning (Danish, Saleh, Andrade, & Bryan, 2017; Yoon et al., 2017). Researchers have found that explicit instruction is necessary to help novices make sense of complex systems (Hmelo-Silver, Marathe & Liu, 2017).

This paper addresses both the call for more attention to learning complex systems in education, as well as the need to research how representational tools can help develop understandings of complex systems. It addresses the question, How do state science education leaders use representations of complex systems to understand their networks and identify leverage points for change? This work is part of a larger research-practice partnership in which state leaders have multiple opportunities to map and analyze the social networks that represent the complex science education systems in their states.

This study employed a mixed methods design which included both qualitative and quantitative data collection and analysis (Leech & Onwuegbuzie, 2009). We analyzed two techniques we constructed that were grounded in Actor Network Theory, which we call state influence charts and actor network maps to capture the ways in which state science education leaders perceive who has power and necessary relationships to make
change in science education. Actor Network Theory (ANT) is concerned with the relations between both material objects and concepts within a network (Latour, 1987, 2005; Sismondo, 2003). ANT posits that there is no unmediated activity within a system, rather, all actions impact and are changed by other actors in the system. Actors—that include human and non-human kinds (e.g., policies operating in a context), have interests that need to be accommodated, and that can be managed and used. Actors work to establish networks that support and constrain activity flowing through that network. ANT involves working to understand the interests of a variety of actors, and translating (both in place and in form) those interests so that actors work together or in agreement (Sismondo, 2003). We use ANT-inspired tools to describe a network as it is, in order to capture an image of the underlying system the produces observable results. Our project is concerned with the political and programmatic work of bringing interests and elements of a network into alignment, for the purposes of increasing coherence and equity in state science education systems. The influence charts and actor network maps capture key decision makers, allies, influential processes, organizations and artifacts, as well as connections between these items in order to make visible the potentially invisible systems of influence in each state.

Description of tools

Network influence maps

The research project associated with this work is organizing networks of science networks, designed to improve equity in science instruction in ways that align with the vision of *A Framework for K-12 Science Education* (National Research Council, 2012). To this aim, the research team is working together with state based science teams to build networks to support coherent and equitable implementation of the new science standards. The project, which began in the Fall of 2016, has included several activities designed to help science coordinators build capacity for change in their state. Some of the activities have included statewide focus groups, the development of rapid assessment tools such as practical measures, and the co-design of professional development resources. The research team has hypothesized that there are many components of coherence in state science education, including the degree to which influential actors within each state are 1) connected and 2) moving in the same direction. State influence charts and actor network maps provided state leaders with an exercise to explore the connectedness and trajectory of important elements of their state systems.

During the summer of 2016 a group of 13 U.S. State Science Coordinators participated in a brief training on how to complete their state influence chart. The purpose of the chart was to guide the science coordinators as they selected a broad team to assist them in implementing a coherent and equitable plan for science teaching and learning in their state. The chart included spaces for coordinators and one to two other team members to list the name and organizations of people who they might want to join their efforts, Table 1. It also included space to write how influential each potential team member was in their state, what aspects of the state science system they had influence over, and how familiar each person was with the *Framework for K-12 Science Education*. Lastly, state coordinators were asked to indicate whether or not the people they listed in their state influence charts knew each other, entering a “1” if the key influencers knew each other, a “0” if they did not and “IDK” if the state coordinator did not know if the potential team members knew each other. The maps were intended to be “living documents” updated as more information was discovered and participants’ roles change. After completing an initial chart, state coordinators met with peers from other states to discuss their charts, the rationale for including people in the chart, and to reflect on what changes they would make to their own charts based on insight from other states. Then state teams revised their influence charts once more during the meeting.

<table>
<thead>
<tr>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
<th>Person 4</th>
<th>Person 5</th>
<th>Person 6</th>
<th>Person 7</th>
<th>Person 8</th>
<th>Person 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>IDK</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Person 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>IDK</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Actor Network Diagrams

State Science Coordinators returned to Boulder for a meeting in Fall 2017. We designed an activity in which members of each state team, including the state science supervisor and one to two other state leaders, completed a hand drawn actor network, Figure 1. Using an interview protocol the research team provided, teams
We brainstormed a list of actors and drew connections between them on paper. We asked state teams to respond to the prompt: “Why does formative assessment in K-12 science classrooms look the way that it does?” Team members were asked to individually brainstorm and then collectively list up to nine actors of their state science systems and to categorize them accordingly: organization, person, process, or artifact. Items categorized as artifacts included the local curriculum or state science standards, while teacher evaluations were an example of an element state teams categorized as a process. Teams were then directed to classify the elements they listed as supportive, a hinderance, or neutral to their efforts to improve state science education. Afterwards, they also entered the information into an online spreadsheet. State teams then used the actor maps to develop detailed action plans related to promoting coherence and equity in state systems. At a later date, the research team will analyze states’ action plans in light of the information they provided in their actor network maps.

The research team used photos of the hand drawn maps to complete any missing elements in the spreadsheets once the activity was over. The spreadsheets were used to make digital maps with an online program called Kumu. Kumu makes a social network map by showing connections between elements, two examples of state maps are below.

**Methods and analysis**

Once the digital version of the actor network maps were created, they were sent to State Science Coordinators in a questionnaire, along with their state influence chart. The questionnaire was sent to the twelve state leads who attended the meeting (1). State science leaders come from a variety of institutions in their state, including state departments of education, local universities, and state boards of education. The questionnaire asked state leaders to compare and contrast their state influence charts and actor network maps, and to share with us what they had learned from these network mapping activities. Questionnaire items included questions such as, “What revisions would you like to make to your Actor Network Map?” and “What stands out to you about the State Influence Chart?” Nine state leaders responded to the questionnaire request. State leader’s survey responses were coded for content to identify broad themes in the text. Some themes that emerged included that the mapping activities were positive experiences for participants, that they helped state leaders identify leverage points of support or opposition in the system, and that state leaders planned to use the mapping activities in conjunction with other practices and activities of the partnership network to plan and make decisions. The original seven categories were combined into three for parsimony: seeing the whole system, planning for the future, and identifying structural gaps.

**Findings**

Supportive, hindering, and neutral actors, processes and organizations

The hand drawn maps provided more information than we had originally anticipated because they allowed participants to create novel connections. For instance, several state coordinators listed organizations that took on more than one role as both supports and hindrances. In all, 15 items were listed as both supporting, hindering
and/or neutral to efforts to improve science education. We did not instruct state coordinators to give elements more than one category because we limited the Kumu spreadsheets to include only one classification. However, state coordinators mapped complexity in their state systems beyond the constraints of the representational system we provided them. They recognized nuance and contingencies in relationships that allowed for organizations to both support and hinder their efforts towards science reform. They described how organizations often have variable interests across individuals, and actors can have shifting interests over time, or variable context-driven interests. We take this to mean that there is significant systems complexity at levels of detail that lie beyond the constraints the activity design imposed on the teams.

![Figure 2](image.png)

*Figure 2*. Example actor network maps. The map on the left represents a somewhat more coherent state with connections between the elements, while the map on the right represents a state with more isolated elements (actants in ANT terms), suggesting incoherence within the state science system. The red elements show misalignment with goals of state leaders, so neither system is coherent with respect to vision for equitable teaching and learning.

**Questionnaire findings**

Our questionnaire findings suggest that representational tools can help develop understandings of complex education systems of which they are part. State science leaders identified several ways that the social actor mapping and state influence chart experiences were positive, with one member calling the activity “reassuring” because she discovered more support in the system than she had originally thought was there. State leaders continued to use ANT-related exercises after the Boulder meeting to organize work in their state. For example, one state leader requested the Kumu spreadsheet so that she could create her own account and lead her team through the data during a meeting. Another science leader completed the mapping activities with others in her state in order to inform decisions regarding leadership for their advisory committee, and to help decide how to redraw regional boundaries for their STEM networks.

Overall, the mapping experiences helped state participants see their state system as an integrated whole, the activities helped them identify incongruencies in state science systems, and the mapping and charting activities helped teams determine their next steps toward improving coherence.

**See the system as a whole**

Respondents felt that the actor network mapping activities helped them see the big picture regarding their state science systems. It was useful for participants to step back to see their entire network as an integrated whole, rather than simply focusing on the parts that they interacted with on a day-to-day basis. One respondent wrote, “It helps me look at all the actors at once and to think about how we can move between them and shift them from hindrance to support.” This state science coordinator reflected on his opportunity to think about the system of science education in his state as a complex whole, rather than a series of discrete, disconnected parts. Similarly, another coordinator stated:

> It was useful to discuss who isn’t part of the conversation and why we haven’t engaged them. It was also critical to see who is influencing the system for better alignment or how [they are] trying to influence a message that may not be in alignment with our vision and mission.
The respondent noticed actors who were a part of the system, and yet had not been engaged in the state science coordinator’s work. The actor network map was useful in identifying parts of the system that their efforts had left untouched. In addition, the state leader identified potential leverage points, by finding actants in their map that may be at odds with the work they were trying to accomplish. These state leaders benefited from the visual representation of the science networks in their state. These representations helped them to locate isolated actants, and started the conversation regarding how better to integrate fragmented sections of the network.

Identify incongruencies
The mapping activities were helpful for illuminating structural gaps in state science systems. One respondent mentioned that there was a disconnect between two major elements on her map. She said, “If we were to look at this, the biggest thing that stands out is that our two major PD opportunities...that get to teachers are completely disconnected.” The recognition of this major discontinuity was facilitated by the visual representation of her state’s science network. Without this activity, the fact that the two primary providers of professional development activities in her state were disconnected—not from the state leader, but from each other—might have gone unnoticed. Building coherence in this network will require connections among all of the major actants, not just between the actants and the science leaders themselves. As mentioned earlier, one participant was surprised to find as much support in her network as she did. Another noted that some of the more influential actants of the network were not having the impact that perhaps they should. The science leader wrote:

It was interesting that not all of the groups/individuals on the influence map are included in the actor map. That is really interesting and we need to consider how to ensure the people who have the most influence are actually acting in our system.

The participant noticed that potentially influential actants were in fact acting in isolation to the rest of the network. This illuminated an incongruence between what science leaders assumed, and what was actually happening upon further inspection. Taken together, the mapping activities provided an opportunity to reflect on whether or not the system was behaving as imagined.

Identify next steps
As a result of the mapping and charting activities, state participants noted some concrete next steps that they could take to build capacity and coherence in the state science systems. For example, participants stated that they needed to expand or extend their networks, and that they would use the charts and maps in order to identify the people to whom they should reach out. One leader noted that he needed to address isolated actants in his actor network map, stating, “I would prefer to see a system that is more tightly clustered. The long tail to the right suggests that I need to do more to bring coherence across influencers.” The same state leader also noted how the maps informed his plan for moving forward in his state. Referring to the state influence chart, he mentioned, “There are a lot of organizations and individuals listed on the document that have not been brought into the work. I need to revisit the document and see how I can draw more people into the work.”

The social network chart and actor map provided a stepping off point for a plan to bring about more coherence in the science network, first by identifying potentially influential people, and then by recognizing that those people had not been recruited into the work. Another state leader responded that the maps helped to think about “specific leverage points” in their system, people and elements that they should directly target for support in their efforts. Similarly, another state science coordinator recognized the need for targeted outreach to enhance her work, stating, “We need more intentional interactions with our messaging to actively seek more champions to elevate the urgency of science education.” The representations of complex systems not only helped state science coordinators visualize actants and clusters of actants of their state science networks, they also provided directions for state coordinators to plan their next course of action.

Limitations
There are some limitations to our activity design. One respondent reported having difficulty enlarging and reading the map, some wanted to list more than the nine elements we constrained them to, and another wanted more time to complete the mapping activity. These are ideas worth exploring. In particular, the research team could assist participants in using their computers’ zoom functions to increase the size of the maps. Given the time constraints at our meeting, we were not able to spend more than 90 minutes on the actor network mapping activity, however, we have continued to work with states who are interested in further developing their maps. Given additional time, the research team could have noted moments of agreement or disagreement as state teams...
negotiated regarding which elements to include in their maps. Our continued work with states can serve as an additional source of validity, as different state team members review the maps and serve as member checks. Despite these limitations, all respondents reported gaining something of use from completing the mapping activities.

Importantly, respondents found the actor network mapping activity to be more informative than the influence chart creation, as the actor network maps included more varied kinds of actants (artifacts, processes, organizations, and people, rather than just people and organizations). However, respondents found that they would need to go back to revise their influence charts after taking into account their actor network maps. Respondents noticed that people who were influential in the state science systems, represented by their actor network maps, were not listed as potential team members in their state influence charts. As the charts were designed to help state leaders compose their science education team, the lack of coordination between the charts and the maps suggests that the two activities can be conducted in concert. Additionally, the actor network mapping activity may be improved in future iterations by encouraging science coordinators to first list all of the influential actants possible before narrowing down the selection to the final nine. This would help to ensure that the nine elements listed were indeed the most influential, rather than simply the first recalled, and the larger list would be a valuable set of data to collect. Participants found it informative to take the entire science network into account in order to build a more effective team.

Discussion
This paper addresses both the call for the use of complex systems in education, as well as the need to research how representational tools and practices can help develop understandings of complex systems. We use a complex system approach to understand state systems of science education, in response to tendencies to imagine science reform as simple and straightforward matters of implementing guidance from research. As one respondent noted, “the system is dynamic and fluxes in one portion of the system impact all other parts of the system.” State science coordinators learned about their networks, the interconnections and disconnections, through using visual representations of their systems. Though state leaders may have had access to all of the information, it was distributed among members and not organized in a way that allowed leaders to view their networks as a whole. The network maps were a helpful tool for mediation that provided the scaffolding necessary to help state leaders to learn about their own complex science systems.

Conclusion
This paper demonstrated a use of complex systems thinking in education and showed that state science leaders can use visual representations of their actor network and state influence charts to better understand their science systems and plan future actions to improve coherence. It helped them identify organizational complexity, new actors, cross-actor tensions, and structural holes in the network to attend to in their improvement work. These interventions are relatively simple and provide affordances for state leaders to think holistically and programmatically about their educational networks.

Endnotes
(1) States included were Arkansas, Iowa, Kansas, New Jersey, Michigan, Minnesota, Oklahoma, Oregon, Pennsylvania, South Dakota, Utah, Washington

References


Teachers’ Values in Co-Design of an Art-Science-Computation Unit

Lila Finch, University of Colorado Boulder, lila.forte@colorado.edu
R. Benjamin Shapiro, University of Colorado Boulder, ben.shapiro@colorado.edu
Franziska Carstens, Westfälische Wilhelms-Universität Münster, franziska.carstens@uni-muenster.de

Abstract: Creating learning environments that bring together the arts, sciences, and computing in education could provide avenues for changing stagnant practices within those disciplines that currently marginalize people with non-normative identities and practices. We examine the co-design process of an interdisciplinary group of teachers crafting a transdisciplinary art-science-computation curriculum unit for their classrooms. Within this context, teachers engaged in richly metarepresentational discussions wherein they applied values and practices from all three constituent fields of study in their design work. We present the results of a qualitative analysis of teacher discourse, showing the wide variety of epistemic criteria marshalled by teachers within design discussions as they connected their ordinarily separate disciplines.

Introduction

Educational researchers have consistently called for more interdisciplinary approaches to education, especially for nurturing stronger, making-oriented synergies between Science, Technology, Engineering, and Math (STEM) and the Arts (STEAM) (Guzdial, 2010; Peppler, 2013; Regalla, 2016), and for valuing a broader set of emotional and epistemic relationships to STEM (Turkle & Papert, 1992). Peppler and Wohlwend (2017) argue that one barrier to equitable participation in STEM is the stagnancy of knowledge development and representation practices, and maker pedagogies that support STEAM learning could broaden participation. For instance, schools expect students in science to produce a small number of canonical representations (e.g. bar graphs) that do little to invite students with experience producing other representations (e.g. art) to apply skills and identities that they already have within science education. They envision the design of nexus learning environments where science-art-computation practices mediate learning and offer new, inclusive affordances for thinking and learning.

Much literature about STEAM describes learning in informal environments (Marallo, 2014; Searle & Kafai, 2015), with some work showing how STEAM can embrace diverse, culturally varying epistemologies (Bang & Medin, 2010). We know little about how nexus learning experiences can unfold within middle and high school classrooms; how new computational-artistic representational practices can be integrated with the existing school curriculum, what kinds of professional development experiences can foster disciplinary teachers’ learning about how to teach using new representational infrastructures (Hall & Jurow, 2015), and what structural barriers within schooling might prevent their implementation.

We present a case study of a multi-school, multi-disciplinary group of teachers co-designing a shared unit that uses artistic and computational representational practices to support investigation and representation of the biology of classroom gardens. We examine how teachers navigate disciplinary values and content expertise in co-designing transdisciplinary projects for their classrooms. We show how metarepresentational discussion (diSessa, 2004) played a crucial role in their making process, including their learning from one another, and shed light on what properties teachers might value in representations that are both scientific and artistic.

Background

On representation as disciplinary practice

Representational infrastructures are the norms for communication about information in disciplines; they are historically embedded, transparent, taken-for-granted, visible only upon breakdown, and stretch across time, space, and social participation structures (Hall & Jurow, 2015; Hall, Stevens, & Torralba, 2002). Often these infrastructures are exclusive, only permitting people to legitimately participate in disciplines in a small set of canonical ways, while excluding people when their backgrounds offer them different representational and epistemic approaches (Bang & Medin, 2010). Nexus learning environments could create room to reimagine representational infrastructures, thereby shaping new forms of induction into disciplinary participation through valuing a multiplicity of ways to contribute. For example e-textiles can change the materiality, the domain of application, and the gendered practices of learning electronics (Peppler & Wohlwend, 2017).

Related to the study of how representational infrastructures define fields is the study of how learning metarepresentational skills can further disciplinary learning. Metarepresentational competence refers to the skills involved in constructing or parsing new representations, comparing sets of representations, and describing the
purpose of a representation (diSessa, 2004). Research on metarepresentation examines both what metarepresentational competence is, and how it develops over time (Danish & Enyedy, 2006; diSessa, Hammer, Sherin, & Kolpakowski, 1991). One strand of that research examines student and teacher thinking about models as representations of scientific phenomena, and in particular, what epistemic criteria they use to evaluate whether a model is a good one (Pluta, Chinn, & Duncan, 2011).

Other work addresses the broader question of how new representational media can support novel ways of learning about and conceptualizing disciplinary content (diSessa, 2001; Greeno & Hall, 1997). Nexus learning environments may provide expansive opportunities for imagining, discussing, and practicing new representational infrastructures, including through the melding of representational techniques across ordinarily separate disciplines. However, there has been little empirical examination of the benefits and drawbacks of such approaches, and we are unaware of any research that investigates how the expansion of teachers’ representational tools and skills can support their development of more representationally and epistemologically ecumenical teaching practices. In this study, we examine teachers’ epistemic criteria within the collaborative construction of artistic and computational representations of scientific phenomena.

On co-design

Collaborative design, or co-design, is a strategy for increasing teacher agency and practical implementation by forming teams of teachers and researchers to design, implement and test new educational innovations (Severance, Penuel, Sumner, & Leary, 2016). Co-design allows for teachers to ensure that reforms amplify their own voices through addressing the contexts, needs, and routines of their practices (Severance et al., 2016).

Storyline authoring is a promising process for science education curriculum co-design (Reiser, 2013). A storyline is a conjecture about how a coherent flow of instruction could be driven by student questions grounded in phenomena that lead to investigations, modeling, and argumentation (Reiser, 2013; Next Generation Science Storylines, n.d.). Thus far, the storylining approach has only been used in contexts where all of the participating teachers are science educators. In this study, we brought together teachers from several different disciplines to collaboratively storyline a nexus unit that blends their expertise in art, science, and computation.

Luminous Science as science-art-computation nexus learning experience

Luminous Science (LS) is an approach developed by the authors that uses art and computation to create dynamic representations of scientific phenomena. This approach was developed to both expand the representational infrastructures of science classrooms, and to create stronger affinities between the values and practices of art, science, and computing education, including artistic inspiration in science. We used a traditional Japanese style of lantern making, Nebuta, and combined it with networked sensors in a garden to create dynamic illuminations in the lantern that are indicators of biochemical phenomena inside the plants. The first author began this project by building her own prototype lantern (Figure 1a) and garden. The Nebuta lanterns use familiar and flexible craft materials (e.g. wire, paper, glue, string, and paint) which allow for many choices in form and aesthetics for visual storytelling. We integrate technology into the lanterns through individually addressable RGB LEDs that are controlled by a programmable microcontroller that communicates wirelessly with another microcontroller and its sensors in a hydroponic garden. Computation, therefore, bridges the artistic and scientific practices of LS.

To tell the story of scientific phenomena in the garden through the lantern, the scientist-artist must make metarepresentational choices about what data to use and then how to visualize the data within the physical shape and dynamic lighting of the lantern. By increasing the scope of possible representational choices, we hope to expand the metarepresentational choice space that students and teachers enjoy. Two scientist-artists making separate lanterns could receive the same data streams but tell quite different stories about the garden through differing choices about how to model and represent the data. Our prototype used data about humidity, temperature, soil moisture, and light as input to a computational model we made about photosynthesis and transpiration. The output of the model is rendered in the body of the lantern via luminous molecules, “sugars” and water, moving through the sculpture, simulating the generation, collection and transportation of molecules in the plant.

We wish to understand how approaches like LS, through creating a nexus of science, art, and computation, could open up science education to a more diverse array of people and practices. As we seek to change science education, we believe it is crucial to study innovations like LS within school, and not just in informal learning spaces. To do so, we must train teachers in the technical facets of the work, work with them to co-design units that they could pilot in their own classroom, and then study the resulting instructional contexts.

Toward that end, we organized a co-design workshop with multi-disciplinary group of teachers, and guided them through a process of transdisciplinary collaboration wherein they made their own LS lanterns and co-wrote unit plans and storylines. Our research questions are:
RQ1: How do teachers discuss the representational values and practices of each other’s disciplines within the co-design process, particularly within transdisciplinary designing and making?

RQ2: What epistemic criteria for science and art do teachers use to make design choices?

Figure 1. a) 9-ft tall prototype Nebuta-style sculpture created by the first author. b-d) 2-4-ft artistic data-driven lanterns made by teachers in the co-design workshop.

Methods

Workshop design

The goal of the workshop was to design a classroom unit around artistically focused data-driven representations of scientific phenomena in gardens that the teachers would deploy in their classrooms. We anticipated that the resulting nexus learning environment designs might necessitate more flexible eventual implementations since an art teacher might have different needs than a science teacher. Therefore, we collaborated with teachers within a framework of co-design to develop storylines for LS that each teacher felt was appropriate for their context.

Within the workshop we asked teachers to 1) make an artistic sensor-data-driven representation of the garden in the form of a lantern and 2) draft a storyline plan for implementation in their classrooms. Our goals for the making activity were to have teachers learn the lantern making process and to facilitate conversations amongst the group of us about how to consider artistic and computational values and practices within science classes, and vice versa. We asked teachers to use a storyline process to develop a sequence of student questions, situated in phenomena, that organize the flow of a unit. Our work here expands on typically mono-subject storyline co-design by bringing together a disciplinarily diverse group of teachers to co-design a nexus storyline.

The workshop ran for five four-hour days. On day 1, we introduced the LS project goals and storylining, sketched lanterns in small groups, and discussed implementation in the classroom. On day 2, we introduced wire frame making and papering, made the wire frames, worked within subject areas on storylines, and the teachers sketched lanterns around a specific garden scenario: Inspired by diSessa et al. (1991), we asked groups to sketch and compare lantern designs that would show the transformation of the garden over time if a water pump failed. On day 3, we introduced data visualization techniques, soldering, and the BBC micro:bit hardware, then groups designed mappings between data and the dynamic lights in their lanterns, then finished wire frames and soldered lights. On day 4, the teachers papered and at least partially painted their lanterns, then shared storylines and planned across schools and subjects. On day 5, we introduced programming sensors and LEDs, and teachers programmed the lights for their lanterns to display data (Figure 1b-d) and worked on their storylines.

Participants

We partnered with six teachers: Hannah (math and computing) and Kate (science), teach in a rural public middle school. Adam (art), Susan (environmental science), Elena (physics), and James (chemistry and earth science) teach in a large public urban high school. Both schools are located in the Rocky Mountain region of the United States. Maggie and Bob, two engineers from a local company that manufactures some of the hardware we used, joined the group for 4 days of the workshop. Bob is a former high school physics teacher, and Maggie is a pre-service teacher. The first author is also a former high school art and chemistry teacher.

Data collection

All workshop activities were recorded using four video and four audio recorders. Two cameras’ locations were fixed and showed an overview of the whole room, but the other two shifted in location based on the evolving focus of group work. The audio recorders supplemented the camera’s audio, and were placed on tables near groups. We took still photos of the groupwork, including photos at the end of each day to document teachers’
products (e.g. plans, lanterns). The teachers wrote their storyline unit outlines and how they would differentiate the unit for their own classes in a shared Google Doc. The teachers wrote responses to a set of daily reflection questions in a Google Form, and we asked additional reflection questions at the end of the week.

Data analysis
Our analysis examined the entire week of video (~70 hrs) and audio data. We took a grounded theory approach (Strauss & Corbin, 1997) to understanding what kinds of discussions about how to represent data occurred within lantern making or unit planning interactions. We performed an exhaustive, low inference search of all the video data to identify metarepresentational discussions, transcribed these, and then grouped the transcribed interactions by their use of common epistemic criteria. The conversations in this paper were transcribed verbatim, and our emphasis was on accuracy of content and sequence rather than on intonation or other markers.

Results
Meta-representational conversations within the nexus of disciplines
Throughout the week, teachers designed and built artistic data-driven lanterns that told an empirical story about a hydroponic garden in our working space. We observed practices and conversations that connected the arts, computing, biology, and physical sciences. The teachers’ interactions with one another were supported by the ways in which the lanterns, garden, and technology provided a context for conversations about what to represent and how to do so using the affordances of the tools at hand. These choices were based on aspects of each discipline that spanned aesthetics, scientific reasoning, and technical possibilities of the hardware and software, showing different areas of expertise, entry points and knowledge sharing. Often, these discussions surfaced epistemic criteria for making design choices. We present a sampling of these metarepresentational discussions below, selected to illustrate the ways in which teachers’ epistemic criteria were raised in design discussions.

Communicative design criteria
Teachers often discussed the aesthetics of their lanterns in terms of communication with a viewer, addressing concerns like how well the form corresponded to the data or the source of the data (the garden), how visually interesting a lantern would be, and the relationship between aesthetic choices and how comprehensible a lantern would be. Some teachers took the stance that the data should drive choices about the aesthetics of their lantern:

Adam: The first thing I’m interested in is finding out is what the data is that we are going to measure.
Kate: That’s going to help to pick the shape of the lantern, for sure.

Other teachers in their group also emphasized correspondences between the visuality of their lantern and the dynamics of the data. Hannah, who, along with Maggie, worked with Adam and Kate, made the case for ensuring that their lantern was visually interesting. Their design had two parts, a box on top that would show incoming data “marching like ants” each time a new data point was received, and the bottom would show a variety of data in three hanging tube structures that would blink or change drastically in color when the values were not in the “normal” range. As they finalized this idea Hannah began to question the visual interestingness of the lantern:

Hannah: [There is] going to be very little of this [lantern] that has any movement to it except for the box which shows the data coming in. The rest of the stuff is just going to be on at a certain level until something goes very wrong, until it’s dying and then it will blink. If I'm a kid that makes me want to kill the plant to see it blink.

Hannah realized that the lantern would always appear the same unless there were something wrong in the plants, making it, in her opinion, somewhat of a boring visualization. So she suggested that they modify the design so that an animation was constantly happening, even if the garden was in normal ranges, and figuring out how students might create more dynamic visuals was a topic that she repeatedly returned to later. However, not all groups seemed to value this correspondence principle; in the next section we describe how Bob and James seemed to make arbitrary decisions about how to connect the form and the data within their design.

Besides being visually appealing, teachers discussed additional epistemic criteria that attended to the distinct but related concerns of simplicity and comprehensibility. Simplicity was valued both in the parsimony of the data presented, as well as in regard to visual complexity. The following interaction illustrates the former:

Bob: So these are three good [types of sensor data] is there anything else we want to add? And I think we might want to keep it simple.
James: I think those three, I think with those three would be the most simple we could make it.

Simplicity of visual representation was sometimes discussed as a way to achieve comprehensibility: paint colors, patterns, structure shapes, or light colors were often chosen based on how easily a viewer would be able to “read” the lantern. Kate expressed a common sentiment of the teachers that overly complex visuals would confuse a viewer when she noted, “But not too much movement though, because then it gets muddied.” instead she emphasized showing a clear, simple story. Other times this visual literacy shaped artistic choices in the paint or lighting patterns, such as when Adam alluded to his expertise in how the light and paint might interact “we could keep the paint pretty neutral because we are using pretty fine color [gradations in the] lights.” The comprehensibility criterion also affected the technical elements of one project. In the following example, Hannah explains to her group why she thinks they should program the lights to be on or off rather than show brightness:

Hannah: We might want to rather than adjusting the brightness of the LEDs we might want to turn more LEDs off or on. Just from playing with the strip of LEDs you can’t tell that much. Like there is a huge range, what I thought was going to be all the way on, medium and off was like medium and all the way on are looking the same to me.

We see here that Hannah noticed a limitation of the effective expressive utility of the lights, specifically that small changes in their brightness are often imperceptible to the human eye (human brightness perception is logarithmic), leading her to recommend a different representation that would still be able to communicate similar information.

Informational design criteria

Within discussions of science and lantern design, teachers mixed concerns for correspondence with other criteria about the need for content coverage and emphasis on “important” scientific information. For instance, one group wanted to add physical motion to their lanterns. In a conversation about when they might use a computer-controlled fan to move paper streamers dangling from their lantern, the group discussed what scientific ideas might be appropriate to represent that way, settling on photosynthesis and respiration:

Maggie: I thought [turning the fan on] was a change in a process? Like when the plant does start to produce O₂ rather than taking oxygen in. [the group looked to Kate to explain the biology]

Adam: Do we have the sensitivity to record that?

Kate: Not being a plant biologist and or even being a biologist, my understanding is that it is always simultaneously happening. You know, we are pretty much constantly taking in oxygen and breathing out CO₂ and plants are pretty much constantly taking in CO₂ and putting out oxygen. That’s my understanding of it. Like at night they are not taking in the same kind of quantity, because they are not really doing photosynthesis in the same way, so it could be a night versus day thing.

Prior to this conversation the group was planning to have the streamers alert the viewer to any large change in any of the data. In thinking about the technological capabilities of the tools in relation to the aesthetic goal of making the streamers move, the group delved into the science behind what scientific processes are dynamic enough to actually map onto the reaction they envisioned. Here the teachers seem to apply a principle of dynamic correspondence: dynamic representational elements should be associated with dynamic scientific phenomena. Although Kate was incorrect about the plant biology, the problem of what to do with the streamers offered an opportunity to the group to consider how representational practices could link science, measurement, and art.

At other times, teachers used concerns about science curriculum content coverage to justify design choices. They began their discussion by focusing on conveying scientific information through the physical shape of their lantern (instantiating the correspondence criterion). They discussed using chemical structures such as buckyballs and DNA, veered off into Viking helmets (tied to the school mascot), and finally settled on mountains. As they proceeded, each story and form they discussed was disjoint from the data and source of that data, the garden, they would be visualizing. The following conversation occurred after they had decided on the structural form of the mountains and were beginning to discuss what biological phenomena they might actually show:

Bob: There could be gnomes… So there are mountain gnomes watching the sensors and that are going to light up, they are going to turn on lights based on what is going on in the plants.

discussion of keeping the representation simple by only using three data points]
Bob: But I think these three would be, so three peaks... we could also talk about ecosystems. That would be good, so if you drive through the mountains, like certain areas are aspen covered, and other areas are pine. [some are] rocky, the top could be the like the white kind of tundra.

James: That depends on how we paint it, one could be the aspen, one could be the pine...

Bob: So ecosystems [lists ecosystems]

James: I think if we took, like the three peaks, even if we can probably combine them... maybe the trees, and that is soil moisture, tundra and snow can be the light and rocky can be the temperature?

Their depiction of mountain ecosystems enhances their lantern’s nominal capacity for content coverage, but the content they add is only loosely connected to observable phenomena within the garden. They seemed to make an arbitrary decision about which data will be shown where on their lantern. We see tensions here between the epistemic criteria of scientific content coverage with both the simplicity and correspondence criteria, as their representation introduces additional visual complexity in order to achieve content coverage, and the visual expression of that content (e.g. snow, trees, rocks) bears no relationship to the particular data they are showing.

We also saw repeated attention to a criterion that Adam called “scientifically interesting data.” For instance, Bob and James’s mountain sculpture design conversation continued as follows:

James: We should be picking [data] on how many, you know how many peaks we are going to have. Three or four is probably a really good... If we wanna stick to three, and choose the most important three, light - definitely.

Bob: Soil moisture like, did you water the plants.

James: So light and soil moisture should be two that should definitely be there.

While some of James and Bob’s choices about how to depict data seem arbitrary, at least one choice of what data to show considered the contextual needs of a human’s care for the garden (“did you water the plants?”).

Transdisciplinary linkages between the nexus disciplines

Even as teachers applied disciplinary epistemic criteria within their metarepresentational work, they also developed transdisciplinary perspectives, drawing unifying connections beyond their individual disciplines (Stember, 1991). They talked with each other about the kinds of questions their disciplines lead them to pose, and discussed how they would apply the perspectives and practices of others’ to expand their own.

Linking ideas and practices between science and art

Though our presentation of some of the epistemic criteria above aligned with a single disciplinary area, teachers’ discussions freely veered between fields. They routinely called attention to the differences between their perspectives, and actively constructed connections between them that seemed to dissolve the distinctions between their fields. For instance, the group below discusses how there is room for difference in implementation of LS between their classrooms, while still allowing for and valuing a convergence of disciplines.

Adam: So, you said something about photosynthesis downstairs and it had me thinking about what’s the transformation of light into energy. Or light into substance actually. Because it takes light and somehow turns it into sugars through the plant’s process... So my questions there was how are things being transformed? Or how do we explore or demonstrate a transformation...

Hannah: So it sounds like your big theme is transformation, where’s mine seems to be more about telling a story with data.

Kate: For me it’s more along the line, you know, how exactly the plant processes and then representing that. The data that we collect.

Researcher: It is possible that we could all have different big questions and ways to introduce it.

Kate: I think that they all kind of go together. At the same time. I mean this data often shows these transformations. And we can do it through the lens of plant processes.

[conversation about whether to teach art history, art techniques and how to teach art]

Adam: From my perspective I would push them into exploring the material in a deeper way, so we would say, okay this is one way of doing it, but what else is the material capable of? How does the ink flow off a brush? How does it respond to being painted on a flat surface versus a dimensional surface.
How does the object take up space? How does it look from all around? These are the sorts of... that I would have.

Hannah: And I think that makes it so much deeper. Like if you are into understanding the materials and the paint, the construction of that thing in a way, as supposed to it being sort of just another version of a poster that tells you what to know about photosynthesis.

From this conversation we see the value that Adam (art teacher) places on materials and capabilities. This conversation seems to influence Hannah (technology teacher) as she starts to connect the deeper understanding and manipulation of materials with better understanding of the science. Similarly, Kate (science teacher) initially focused solely on the plant processes, but as the conversation progressed she saw all the big ideas as inherently tied together. Through these discussions new types of questions teachers might ask their students or the way they view their content could be reformulated in ways that would not be traditional for their specific classroom.

**Learning computing through attention to artistic and scientific criteria**

At times, their aesthetic or scientific aims motivated teachers to learn computing and electronics skills, such as when Adam identified the electronics skills he would need to learn to implement a design goal of the group, by asking how to “arrange the circuits”. In another episode, James and Bob wanted a string of lights to blink at a speed proportional to light sensor data. They had an aesthetic desire to depict the garden data a certain way, but to do that James, who had never programmed before, needed to learn about variables, how to manipulate incoming data, and how to create an animation of moving lights. In this episode, James programs and Bob helps him:

James: So now if we set the light level [sets the incoming light data to a variable named “light level”]

James: So now it will go faster if [the incoming light] is brighter? [finishes the program and they test it]

Bob: So now let’s add more light. [physically shines a flashlight at the sensor; lantern responds by having a slower movement of the lights in a chasing motion]

James: Right, so a longer pause because it’s a bigger number, right that makes sense, so it actually blinks slower.

Their representational choice originally had been that more light would make the lights go faster, but because they used a linear mapping from the sensor data to pause time, James found it had the inverse effect. The overlapping of art, science, and technology practices within this project allowed James to learn not just about the science, but created a meaningful context to learn programming from Bob in order to achieve the artistic representational ends he desired, and to observe the cause of an unintended relationship between data and their visual representation.

**Discussion and conclusion**

Through this weeklong transdisciplinary teacher workshop, we examined how co-design of a nexus learning environment could support teachers’ explorations of representational practices and values from each other’s disciplines. As they prototyped lanterns to visualize scientific data, they learned to use new representational forms (e.g. Nebuta lanterns, computer code) and tools (e.g. programmable micro-controllers).

Metarepresentational competence has generally been valued in the literature as mastery over a set of skills that support students’ science learning. In some of that work, students are invited to invent their own representations, before being pushed to work with canonical representations (diSessa et al., 1991). In contrast, our co-design processes envisioned learning experiences that value students’ artistic representational inventions in and of themselves. Further, teachers’ lantern construction work functioned as a nexus learning environment for them, with metarepresentational discourse being central to the teachers’ collaborations with one another. In particular, the question of how to present a meaningful representation of data was a critical thread that connected disciplinary practices and values from the arts, computing, and sciences. The project allowed for multiple ways to be recognized as expert within the teachers’ collaborations, illustrating how nexus learning environments can create room to value a more diverse set of skills and perspectives within STEM education, and change what “counts” as a science representation, or even as a way of “doing science” or “doing art.” We note that teachers’ discussions of criteria for their lanterns seemed to address a larger terrain than prior research on epistemic criteria has attended to; while we can conceive of the lanterns as models of biological phenomena (as in Pluta, Chinn, & Duncan, 2011), we, the researchers, and the teachers both valued elements of them that went beyond representing scientific processes (e.g., Bob and James’ gnomes, and the face of the “tree spirit” in our own lantern).

A limitation of the workshop was that it only included one active art teacher, so each group did not have as much direct access to the perspectives and practices of an expert art teacher; the metarepresentational talk within the most diverse group (with CS, art, and science teachers) seemed especially rich to us. Adam’s expertise...
on art and art teaching was called upon by other groups. Additionally, groups tended to have only a single teacher from each discipline, so there were times when scientific misinformation went unchallenged, something less likely to occur within mono-discipline co-design. An intriguing result of the group’s work together was the production of several storylines with overlapping but divergent narratives. The teachers discussed values and practices from each other’s disciplines, noting affordances for their own work, which led to overlaps in storylines; but also noting challenges due to differing contexts (e.g. schools, concepts) which led to differentiations in storylines. We will address these convergences and divergences during the co-design process in future papers.

As this was a study of teachers’ collaborating in co-design, and not a classroom implementation, we do not yet know how this nexus approach would create opportunities for more inclusive student participation in science. We are optimistic that the generousness with which the teachers embraced each other’s perspectives within the co-design would translate to their classroom practices. The teachers who participated in this study are currently implementing LS in their classrooms; in a future paper we will examine how these implementations realize our aims of expanding the kinds of values and practices that can exist simultaneously in the classroom, and how doing so creates space for teachers to apply their disciplinary expertise in support of learning experiences for students that embrace a variety of identities, values, and practices for learning.

References

Acknowledgements
We thank Oracle and the National Science Foundation (awards 1227530 and 1562040) for supporting this work.
Indexes
Author Index

Pages 1-670: Volume 1
Pages 671-1342: Volume 2
Pages 1343-1888: Volume 3

Abrahamson, Dor, 1089, 1243, 1405
Acharya, Arup, 1391
Acosta, Alisa, 432, 1259
Adams, Deanne, 1259
Adams, Jennifer, 1320
Adams, Joshua, 953
Addeh, Afe, 64
Agarwal, Priyanka, 832
Ahn, June, 705, 1673
Ainsworth, Shaaron, 941, 969
Aleven, Vincent, 104, 304, 689, 1227, 1541
Allen, Shawndra, 1312
Alozie, Nonye, 1211
Anderson, Alice, 981
Anderson, Charles W., 1235
Anderson, Eleanor, 1219
Anderson, Emma, 224, 1275
Anderson, Janice L., 1703
Anderson, Richard C., 977
Anderson, Robin, 1511
Anderson, Theresa Dirndorfer, 941
Anderson-Levitt, Kathryn, 713
Andrade, Alejandro, 1243
Andrews, Chelsea J., 1649
Anton, Gabriella, 1617
Applebaum, Lauren, 1473
Arastoopour Irgens, Golnaz, 871
Archibeque, Benjamín, 1081
Arend, Béatrice, 1459
Arias, Anna Maria, 1357
Arukovic, Selma, 1583
Asaridou, Salomi S., 1767
Asino, Tutaleni I., 1571
Askari, Emilia, 344
Asterhan, Christa, 400
Atarashi, Erika, 1525
Atit, Kimi, 961
Atkins Elliott, Leslie, 1679
Avry, Sunny, 184
Av-Shalom, Na'ama, 768
Bae, Haesol, 1483
Bain, Connor, 871
Baker, Ryan, 1385
Baker, Sara, 1101
Baker-Doyle, Kira, 1235
Bakker, Arthur, 1243
Bang, Megan, 1187, 1283, 1312
Banks, Tarrence, 1165
Barab, Sasha, 1328
Baradaran, Nilou, 1153
Baram-Tsabari, Ayelet, 416, 1467
Barany, Amanda, 1663
Barber-Lester, Kelly, 1703
Barbieri, Christina, 1581
Bar-El, David, 1759
Barma, Sylvie, 1153, 1320
Barnes, Jackie, 1259
Barron, Brigid, 1187
Bartell, Tonya, 1141
Barth-Cohen, Lauren, 1599, 1679
Bazizilai, Sarit, 1129, 1251
Bastani, Reyhaneh, 1743
Basu, Satabdi, 384, 925, 1304, 1379
Bateman, Kathryn, 985
Bayeck, Rebecca Yvonne, 1611
Baylor, Amy L., 1843
Beck, Stephanie, 1153
Becker, Sandra, 1105
Bedell, Kristin Dana, 1259
Behrendt, Lena, 1061
Bell, Amanda M., 917, 1701
Bell, Amani, 1713
Bell, Philip, 656, 1187, 1729
Ben-David Kolikant, Yifat, 1535
Bender, Sophia, 1605
Benedict-Chambers, Amanda, 1357
Ben-Horin, Hava, 1259
Ben-Zvi, Dani, 376
Berg, Yngvar, 1577
Berland, Leema, 1251, 1433
Berland, Matthew, 1745, 1487, 1745
Bernstein, Debra, 1413
Betser, Sagit, 847, 1455, 1661
Bhaduri, Srinjita, 1513
Bhattacharya, Devarati, 1397
Bian, Yulong, 1371
Bienkowski, Marie, 1377
Bild, David, 1312
Binder, Torsten, 528
Biswas, Gautam, 384, 1691
Blikstein, Paulo, 544, 800, 839, 1275, 1531, 1685, 1749
Boardman, Alison, 1669
Bod, Petra, 1365
Bodemer, Daniel, 456, 1389, 1457, 1555
Bohrmann-Linde, Claudia, 937
<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrigan, Seth</td>
<td>512</td>
</tr>
<tr>
<td>Costa, Stacy</td>
<td>1719</td>
</tr>
<tr>
<td>Courshon, Caitlin</td>
<td>1747</td>
</tr>
<tr>
<td>Craig, Kalani</td>
<td>1483</td>
</tr>
<tr>
<td>Cress, Ulrike</td>
<td>1551</td>
</tr>
<tr>
<td>Crowley, Kevin</td>
<td>1641</td>
</tr>
<tr>
<td>Cukurova, Mutlu</td>
<td>681, 1549</td>
</tr>
<tr>
<td>Cullicot, Catherine E.</td>
<td>1479</td>
</tr>
<tr>
<td>Cun, Aijuan</td>
<td>1145</td>
</tr>
<tr>
<td>Curwood, Jen Scott</td>
<td>1713</td>
</tr>
<tr>
<td>Cutumisu, Maria</td>
<td>1537</td>
</tr>
<tr>
<td>Dabholkar, Sugat</td>
<td>1617</td>
</tr>
<tr>
<td>Dado, Marielle</td>
<td>456</td>
</tr>
<tr>
<td>Dagan, Efrat</td>
<td>949</td>
</tr>
<tr>
<td>Dahn, Maggie</td>
<td>72, 1243, 1431</td>
</tr>
<tr>
<td>Damsa, Crina</td>
<td>1577</td>
</tr>
<tr>
<td>Danahy, Ethan</td>
<td>1413</td>
</tr>
<tr>
<td>Dance, Sequoia Lynn</td>
<td>697</td>
</tr>
<tr>
<td>D'Angelo, Cynthia</td>
<td>1687</td>
</tr>
<tr>
<td>Daniel, Shannon M.</td>
<td>1761</td>
</tr>
<tr>
<td>Danish, Joshua</td>
<td>72, 512, 584, 1077, 1243, 1275, 1431, 1723, 1763</td>
</tr>
<tr>
<td>Dasgupta, Chandan</td>
<td>1359, 1529, 1852</td>
</tr>
<tr>
<td>Datu, Emmanuella</td>
<td>977</td>
</tr>
<tr>
<td>Datye, Abhaya</td>
<td>1427</td>
</tr>
<tr>
<td>Davis, Bria</td>
<td>1243, 1275, 1763</td>
</tr>
<tr>
<td>Davis, Darrel</td>
<td>1653</td>
</tr>
<tr>
<td>Davis, Pryce</td>
<td>1367, 1603</td>
</tr>
<tr>
<td>Davis, Richard Lee</td>
<td>800</td>
</tr>
<tr>
<td>Day, Natalie</td>
<td>1101</td>
</tr>
<tr>
<td>De Backer, Liesje</td>
<td>1296</td>
</tr>
<tr>
<td>de Freitas, Elizabeth</td>
<td>1195</td>
</tr>
<tr>
<td>de Laat, Maarten</td>
<td>1235</td>
</tr>
<tr>
<td>de los Santos, Elizabeth</td>
<td>1235</td>
</tr>
<tr>
<td>Deitrick, Elise</td>
<td>1121</td>
</tr>
<tr>
<td>DeLaRosa, Mia</td>
<td>887</td>
</tr>
<tr>
<td>Delgado, Cesar</td>
<td>1607</td>
</tr>
<tr>
<td>DeLiem, David</td>
<td>1085, 1405</td>
</tr>
<tr>
<td>DellaVecchia, Gabriel</td>
<td>1089</td>
</tr>
<tr>
<td>Demir-Lira, Ozlem Ece</td>
<td>1767</td>
</tr>
<tr>
<td>Denton, Rebecca</td>
<td>1585</td>
</tr>
<tr>
<td>Depaepe, Fien</td>
<td>128, 408</td>
</tr>
<tr>
<td>Derry, Sharon</td>
<td>1703</td>
</tr>
<tr>
<td>Desmet, Piet</td>
<td>128</td>
</tr>
<tr>
<td>DeVane, Benjamin</td>
<td>1349, 1747</td>
</tr>
<tr>
<td>Di Tosto, Gennaro</td>
<td>905</td>
</tr>
<tr>
<td>Diana, Nicholas</td>
<td>1377</td>
</tr>
<tr>
<td>Diaz Caceres, Zulay</td>
<td>1105</td>
</tr>
<tr>
<td>Dich, Yong Li</td>
<td>608</td>
</tr>
<tr>
<td>Dickler, Rachel</td>
<td>1465</td>
</tr>
<tr>
<td>Diera, Claudia</td>
<td>713</td>
</tr>
<tr>
<td>Dietmeier, Jeremy</td>
<td>1349, 1747</td>
</tr>
<tr>
<td>Dietz, Brianna</td>
<td>1765</td>
</tr>
<tr>
<td>Dillenbourg, Pierre</td>
<td>640</td>
</tr>
<tr>
<td>Ding, Lu</td>
<td>953</td>
</tr>
<tr>
<td>Dixon, Colin G.</td>
<td>1455</td>
</tr>
<tr>
<td>DMello, Sidney</td>
<td>1385</td>
</tr>
<tr>
<td>Doll, Jenifer</td>
<td>913</td>
</tr>
<tr>
<td>Donaldson, Jonan Phillip</td>
<td>1419, 1663</td>
</tr>
<tr>
<td>Dornfeld Tissenbaum, Catherine Louise</td>
<td>240, 1069, 1595</td>
</tr>
<tr>
<td>Doroudi, Shayan</td>
<td>1267, 1886</td>
</tr>
<tr>
<td>Doubler, Susan</td>
<td>1005</td>
</tr>
<tr>
<td>Dowell, Nia</td>
<td>1421, 1815</td>
</tr>
<tr>
<td>Dragnic-Cindrie, Dalila</td>
<td>1631</td>
</tr>
<tr>
<td>Dresel, Markus</td>
<td>1296, 1561</td>
</tr>
<tr>
<td>Dubovi, Ilana</td>
<td>949, 1057</td>
</tr>
<tr>
<td>Duesbery, Luke</td>
<td>1639</td>
</tr>
<tr>
<td>Dugdale, Michael</td>
<td>1651</td>
</tr>
<tr>
<td>Duncan, Ravit Golan</td>
<td>576, 768, 808, 1251</td>
</tr>
<tr>
<td>Duvall, Matthew</td>
<td>1579</td>
</tr>
<tr>
<td>Dvir, Michal</td>
<td>376</td>
</tr>
<tr>
<td>Dyer, Elizabeth B.</td>
<td>1387, 1854</td>
</tr>
<tr>
<td>Eagan, Brendan R.</td>
<td>1625</td>
</tr>
<tr>
<td>Eagle, Michael</td>
<td>1377</td>
</tr>
<tr>
<td>Easley, Kathleen M.</td>
<td>1629</td>
</tr>
<tr>
<td>Easterday, Matthew</td>
<td>9, 352</td>
</tr>
<tr>
<td>Eberle, Julia</td>
<td>1583</td>
</tr>
<tr>
<td>Edwards, Kelsey</td>
<td>25</td>
</tr>
<tr>
<td>Ehrenfeld, Nadav</td>
<td>1627</td>
</tr>
<tr>
<td>Elby, Andrew</td>
<td>1477</td>
</tr>
<tr>
<td>Elen, Jan</td>
<td>128</td>
</tr>
<tr>
<td>Eliason, Chad</td>
<td>1361</td>
</tr>
<tr>
<td>El-Moslimany, Hebah</td>
<td>768</td>
</tr>
<tr>
<td>Enyedy, Noel</td>
<td>72, 1243, 1275, 1431, 1763</td>
</tr>
<tr>
<td>Erickson, Tim</td>
<td>1383</td>
</tr>
<tr>
<td>Erkens, Melanie</td>
<td>1555</td>
</tr>
<tr>
<td>Essmiller, Kathy</td>
<td>1571</td>
</tr>
<tr>
<td>Eylon, Bat-Sheva</td>
<td>1259</td>
</tr>
<tr>
<td>Feinstein, Noah Weeth</td>
<td>1433</td>
</tr>
<tr>
<td>Feldon, David F.</td>
<td>560</td>
</tr>
<tr>
<td>Feng, Mingyu</td>
<td></td>
</tr>
<tr>
<td>Feng, Xueqi</td>
<td>1737</td>
</tr>
<tr>
<td>Fernandez, Cassia</td>
<td>1749</td>
</tr>
<tr>
<td>Ferrara, Francesca</td>
<td>1195</td>
</tr>
<tr>
<td>Ferrari, Giulia</td>
<td>1195</td>
</tr>
<tr>
<td>Fick, Sarah J.</td>
<td>1357</td>
</tr>
<tr>
<td>Fields, Deborah</td>
<td>721</td>
</tr>
<tr>
<td>Finch, Lila</td>
<td>663</td>
</tr>
<tr>
<td>Finzer, William</td>
<td>1383</td>
</tr>
<tr>
<td>Fisher, Douglas</td>
<td>152</td>
</tr>
<tr>
<td>Fitzgerald, Miranda S.</td>
<td>616, 1089, 1629</td>
</tr>
<tr>
<td>Flanagan-Hall, Grace Ann</td>
<td>961</td>
</tr>
<tr>
<td>Flood, Virginia J.</td>
<td>1085, 1405</td>
</tr>
<tr>
<td>Flores, Yuliana</td>
<td>737</td>
</tr>
<tr>
<td>Floryan, Mark</td>
<td>1359</td>
</tr>
<tr>
<td>Forbes, Cory</td>
<td>1397</td>
</tr>
<tr>
<td>Fortus, David</td>
<td>1211, 1369</td>
</tr>
<tr>
<td>Fowler, Teresa</td>
<td>1471</td>
</tr>
<tr>
<td>Francois, Annmarie</td>
<td>713</td>
</tr>
<tr>
<td>Frank, Kenneth A.</td>
<td>1235</td>
</tr>
<tr>
<td>Franke, Baker E.</td>
<td>328</td>
</tr>
<tr>
<td>Franklin, Scott</td>
<td>1081, 1623</td>
</tr>
<tr>
<td>Fratamico, Lauren</td>
<td>1409</td>
</tr>
</tbody>
</table>
Friend, Michelle, 800
Fritzsche, Eva S., 1601
Froehlich, Jon, 64, 1275
Fu, Chuxin, 1435
Fuhrmann, Tamar, 1685, 1749
Furberg, Anniken, 989
Furtak, Erin Marie, 520, 528, 1211
Gabai, Joshua, 1745
Gadgil, Soniya, 536
Gal, Kobi, 1227
Galvin, Sarah, 344
Gane, Brian Douglas, 1211, 1495, 1521
Gao, Dandan, 784
Garcia, Antero, 1669
Garcia, Brittany, 1699
Garder, Joshua, 1835
Gardner, Stephanie, 1251
Gargroetzi, Emma, 745
Garza, Veronica, 160
Genz, Florian, 1081
George, MariAnne, 1521
Georgen, Chris, 1243, 1275, 1882
Georgiou, Yiannis, 392
Gerard, Libby, 816, 1259
Gerber, Elizabeth, 9, 352
Giannakos, Michail, 640
Gobert, Janice, 1465
Godley, Amanda, 296
Goede, Megan, 981
Gold, Anne U., 1715
Goldin-Meadow, Susan, 1767
Goldman, Shelley, 1187, 1659
Goldman, Susan R., 440, 1521
Goldwater, Micah, 200
Gomoll, Andrea Sarah, 1407, 1563, 1609
Gonzales, Wendy, 1312
González, Lydia Marion, 1537
Goodwin, Amanda, 152
Goodyear, Peter, 1219
Gorson, Jamie, 9, 352
Gould, Deena, 1771
Gould, Mary, 1501
Gouvea, Julia Svoboda, 33, 1137
Gravel, Brian, 921
Graville, Cynthia, 1501
Greenberg, Day, 1320
Greene, Jeffrey Alan, 1631
Greenhow, Christine, 344
Grenshl, Jana, 1473, 1497
Greisel, Martin, 1296, 1561
Gresalfi, Melissa, 1335, 1701
Griffin, Jane, 1397
Grinath, Anna Strimaitis, 1451
Grover, Shuch, 384, 925, 1304, 1377, 1447
Gu, Xiaoping, 1435
Gulz, Agneta, 1463
Gupta, Ayush, 1219
Gutiérrez, José, 1021
Guyevskey, Victoria, 1335
Haganah, Sara, 1320
Haines, Cara, 1169
Hakkarainen, Kai, 248, 1379, 1417
Hall, Allison A., 440, 1485
Hall, Kevin, 871
Hall, Rogers, 1195, 1312
Hallinen, Nicole, 1647
Halverson, Erica, 921
Halverson, Richard, 1203
Hamada, Rhonya A., 1687
Hämäläinen, Raija, 1296
Hamaoui, Marie, 1449
Hamilton, Megan M., 1739
Han, Insook, 1689
Han, Shenshen, 1077
Hand, Victoria, 1141
Härkkälä, Veeti, 1379
Harrer, Benedikt Walter, 1405
Harris, Christopher J., 384, 1211
Harrison, Emily Jean, 816
Hartmann, Christian, 1125, 1681
Hart, Martha, 536
Haslbeck, Heidi, 1601
Hathaway, Kennedy, 1765
Hatten-Roberts, Toni M., 1423
Headrick Taylor, Katie, 752
Heffernan, Neil T., 1267
Heimbuch, Sven, 1387
Heller, Joan I., 1383
Hennessy Elliott, Colin, 48, 1041
Henrichson, Clayton, 1585
Henry, Jessica, 1489
Henson, Kate, 528, 1597
Heo, Gyeong Mi, 1533
Herbel-Eisenmann, Beth, 1141
Herman, Phillip, 632
Hernly, Kenna, 705
Herrenkohl, Leslie Rupert, 1290
Hershman, Anneli, 1409
Hew, Khe Foon, 1533
Heyd-Metzuyanim, Einat, 1627
Hickey, Daniel T., 792
Hicks, David, 136
Hills, Thomas, 1267
Hilppö, Jaakko, 1203
Hinojosa, Leihanna, 120, 1499, 1517
Hjorth, Arthur, 1219
Hladik, Stephanie, 1320
Hmelo Silver, Cindy, 1407, 1483, 1563, 1609, 1831
Hoadley, Christopher, 1671
Hochgreb-Haegele, Tatiana, 1749
Hod, Yotam, 376, 1033
<table>
<thead>
<tr>
<th>Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pham, Duy</td>
<td>1503</td>
</tr>
<tr>
<td>Pham, Josephine</td>
<td>713</td>
</tr>
<tr>
<td>Philip, Thomas M.</td>
<td>1219, 1283</td>
</tr>
<tr>
<td>Phillips Galloway, Emily</td>
<td>1693</td>
</tr>
<tr>
<td>Phillips, Abigail Leigh</td>
<td>320</td>
</tr>
<tr>
<td>Phillips, Anna McLean</td>
<td>192</td>
</tr>
<tr>
<td>Phillips, Nathan C.</td>
<td>1312</td>
</tr>
<tr>
<td>Pinkard, Nichole</td>
<td>1187</td>
</tr>
<tr>
<td>Pitt, Caroline</td>
<td>705</td>
</tr>
<tr>
<td>Plummer, Julia</td>
<td>929</td>
</tr>
<tr>
<td>Pollack, Sarah</td>
<td>1535</td>
</tr>
<tr>
<td>Polman, Joseph</td>
<td>1001, 1290, 1499, 1517, 1669</td>
</tr>
<tr>
<td>Pon-Barry, Heather</td>
<td>737</td>
</tr>
<tr>
<td>Popov, Vitaliy</td>
<td>160, 855</td>
</tr>
<tr>
<td>Powell, Sandy</td>
<td>1157</td>
</tr>
<tr>
<td>Pozen, Rebecca</td>
<td>1625</td>
</tr>
<tr>
<td>Price, Aaron</td>
<td>1473, 1475, 1495, 1497</td>
</tr>
<tr>
<td>Price-Dennis, Detra</td>
<td>1113</td>
</tr>
<tr>
<td>Proctor, Chris</td>
<td>544, 800</td>
</tr>
<tr>
<td>Prudent, Miranda E.</td>
<td>1773</td>
</tr>
<tr>
<td>Prusak, Naomi</td>
<td>1227</td>
</tr>
<tr>
<td>Puntambekar, Sadhana</td>
<td>240, 280, 1595</td>
</tr>
<tr>
<td>Puttick, Gillian</td>
<td>1413</td>
</tr>
<tr>
<td>Qi, Guowen</td>
<td>1371</td>
</tr>
<tr>
<td>Qiu, Chen</td>
<td>1427</td>
</tr>
<tr>
<td>Quinn, Kimberly A.</td>
<td>1473</td>
</tr>
<tr>
<td>Quintana, Rebecca M.</td>
<td>1415</td>
</tr>
<tr>
<td>Radinsky, Josh</td>
<td>1290, 1312</td>
</tr>
<tr>
<td>Radke, Sarah C.</td>
<td>48, 1093, 1335</td>
</tr>
<tr>
<td>Raes, Annelies</td>
<td>408</td>
</tr>
<tr>
<td>Rahimi, Seyadahmad</td>
<td>1385</td>
</tr>
<tr>
<td>Rahm, Irene</td>
<td>1320</td>
</tr>
<tr>
<td>Rajala, Antti</td>
<td>1203</td>
</tr>
<tr>
<td>Rajendran, Ramkumar</td>
<td>1691</td>
</tr>
<tr>
<td>Rambo-Hernandez, Karen</td>
<td>1765</td>
</tr>
<tr>
<td>Ramey, Kay E.</td>
<td>168, 961, 1203</td>
</tr>
<tr>
<td>Rau, Martina</td>
<td>264, 272, 504</td>
</tr>
<tr>
<td>Recker, Mimi</td>
<td>320</td>
</tr>
<tr>
<td>Rees Lewis, Daniel</td>
<td>9, 352</td>
</tr>
<tr>
<td>Reich, Justin</td>
<td>648</td>
</tr>
<tr>
<td>Reichsman, Frieda</td>
<td>1383</td>
</tr>
<tr>
<td>Reilly, Joseph</td>
<td>448, 608</td>
</tr>
<tr>
<td>Reimann, Peter</td>
<td>1328</td>
</tr>
<tr>
<td>Reiser, Brian</td>
<td>25</td>
</tr>
<tr>
<td>Renken, Maggie</td>
<td>729, 1505</td>
</tr>
<tr>
<td>Resnick, Lauren</td>
<td>1769</td>
</tr>
<tr>
<td>Resnick, Maya S.</td>
<td>400</td>
</tr>
<tr>
<td>Ricca, Bernard</td>
<td>1725</td>
</tr>
<tr>
<td>Richard, Gabriela</td>
<td>1755</td>
</tr>
<tr>
<td>Richards, Jennifer</td>
<td>1017</td>
</tr>
<tr>
<td>Richey, J. Elizabeth</td>
<td>1653</td>
</tr>
<tr>
<td>Ricketts, Amy</td>
<td>929</td>
</tr>
<tr>
<td>Rico-Beck, Laura</td>
<td>1495</td>
</tr>
<tr>
<td>Riedel-Kruse, Ingmar</td>
<td>839, 1275</td>
</tr>
<tr>
<td>Riedy, Robbin</td>
<td>656</td>
</tr>
<tr>
<td>Riesbeck, Christopher</td>
<td>9</td>
</tr>
<tr>
<td>Riikonen, Sini</td>
<td>248</td>
</tr>
<tr>
<td>Rish, Ryan</td>
<td>1145</td>
</tr>
<tr>
<td>Ristvey, John</td>
<td>1513</td>
</tr>
<tr>
<td>Roberts, Jessica</td>
<td>1641</td>
</tr>
<tr>
<td>Robinson, Kevin</td>
<td>648</td>
</tr>
<tr>
<td>Rogge, Carson</td>
<td>871</td>
</tr>
<tr>
<td>Romine, William</td>
<td>808</td>
</tr>
<tr>
<td>Roque, Ricarose</td>
<td>592, 1187</td>
</tr>
<tr>
<td>Roschelle, Jeremy</td>
<td>632</td>
</tr>
<tr>
<td>Rosé, Carolyn</td>
<td>296, 472, 1267</td>
</tr>
<tr>
<td>Rosebery, Ann</td>
<td>1283</td>
</tr>
<tr>
<td>Rosier, Sorensen</td>
<td>1709</td>
</tr>
<tr>
<td>Ross, Jen</td>
<td>1713</td>
</tr>
<tr>
<td>Rothschild, Stefan</td>
<td>1105</td>
</tr>
<tr>
<td>Rourke, Liam</td>
<td>1615</td>
</tr>
<tr>
<td>Roy, Dan</td>
<td>648</td>
</tr>
<tr>
<td>Roy, Deb</td>
<td>1409</td>
</tr>
<tr>
<td>Royse, Emily</td>
<td>997</td>
</tr>
<tr>
<td>Ruis, Andrew</td>
<td>1625</td>
</tr>
<tr>
<td>Rummel, Nikol</td>
<td>689, 1045, 1125, 1227, 1541, 1583, 1681</td>
</tr>
<tr>
<td>Runburg, Madlyn</td>
<td>1683</td>
</tr>
<tr>
<td>Russ, Rosemary</td>
<td>1433</td>
</tr>
<tr>
<td>Russell, Randy</td>
<td>1513</td>
</tr>
<tr>
<td>Rutstein, Daisy</td>
<td>1304</td>
</tr>
<tr>
<td>Ryoo, Kihyun “Kelly”</td>
<td>1259</td>
</tr>
<tr>
<td>Ryu, Minjung</td>
<td>1761</td>
</tr>
<tr>
<td>Ryu, Suna</td>
<td>1711</td>
</tr>
<tr>
<td>Saarinen, Auli</td>
<td>1417</td>
</tr>
<tr>
<td>Sabanovic, Selma</td>
<td>1407, 1563, 1609</td>
</tr>
<tr>
<td>Sagy, Ornit</td>
<td>1259</td>
</tr>
<tr>
<td>Saito, Moegi</td>
<td>1525, 1807</td>
</tr>
<tr>
<td>Saiyera, Tashire</td>
<td>1537</td>
</tr>
<tr>
<td>Saleh, Asmalina</td>
<td>1243, 1831</td>
</tr>
<tr>
<td>Samarapungavan, Ala</td>
<td>1251</td>
</tr>
<tr>
<td>Sandoval, William A.</td>
<td>760, 1251</td>
</tr>
<tr>
<td>Santo, Rafi</td>
<td>1235, 1671, 1673</td>
</tr>
<tr>
<td>Sarmiento, Juan Pablo</td>
<td>1259, 1673</td>
</tr>
<tr>
<td>Sayre, Eleanor C.</td>
<td>957, 1081, 1623</td>
</tr>
<tr>
<td>Scardamalia, Marlene</td>
<td>1719</td>
</tr>
<tr>
<td>Schall, Max</td>
<td>937</td>
</tr>
<tr>
<td>Scheiter, Katharina</td>
<td>937</td>
</tr>
<tr>
<td>Schimpf, Corey</td>
<td>1359</td>
</tr>
<tr>
<td>Schloss, Bob</td>
<td>1391</td>
</tr>
<tr>
<td>Schmidt, Ani</td>
<td>1312</td>
</tr>
<tr>
<td>Schnaubert, Lenka</td>
<td>1457</td>
</tr>
<tr>
<td>Schneider, Bertrand</td>
<td>448, 608</td>
</tr>
<tr>
<td>Schneider, Leslie</td>
<td>1489</td>
</tr>
<tr>
<td>Schneitzler, Malin Kimberley</td>
<td>1555</td>
</tr>
<tr>
<td>Schönfeld, Tim</td>
<td>1583</td>
</tr>
<tr>
<td>Schreurs, Bieke</td>
<td>1235</td>
</tr>
<tr>
<td>Schunn, Chris</td>
<td>568, 1009</td>
</tr>
<tr>
<td>Schwartz, Daniel L.</td>
<td>1537</td>
</tr>
<tr>
<td>Schwarz, Baruch</td>
<td>1227</td>
</tr>
<tr>
<td>Scipio, Deana A.</td>
<td>192, 1866</td>
</tr>
<tr>
<td>Scroggins, Ashley</td>
<td>1141</td>
</tr>
<tr>
<td>Searcy, John</td>
<td>1165</td>
</tr>
<tr>
<td>Searle, Kristin A.</td>
<td>560, 697</td>
</tr>
<tr>
<td>Name</td>
<td>Page Numbers</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Thompson, Meredith</td>
<td>648</td>
</tr>
<tr>
<td>Thompson, Naomi</td>
<td>512, 584, 1077</td>
</tr>
<tr>
<td>Thompson, Penny</td>
<td>1571</td>
</tr>
<tr>
<td>Tierney, Gavin</td>
<td>1613, 1619, 1753</td>
</tr>
<tr>
<td>Tissenbaum, Mike</td>
<td>1304, 1657</td>
</tr>
<tr>
<td>Todd, Amber</td>
<td>808</td>
</tr>
<tr>
<td>Tofel-Greul, Colby</td>
<td>560</td>
</tr>
<tr>
<td>Tolar, Erin</td>
<td>1407, 1563, 1609</td>
</tr>
<tr>
<td>Tong, Yurui</td>
<td>312</td>
</tr>
<tr>
<td>Tong, Yuyao</td>
<td>879, 1775</td>
</tr>
<tr>
<td>Toprani, Dhwani</td>
<td>1573</td>
</tr>
<tr>
<td>Torralba, Juan</td>
<td>1449</td>
</tr>
<tr>
<td>Totten, Tatia B.</td>
<td>1573</td>
</tr>
<tr>
<td>Touliou, Ferdous</td>
<td>1320</td>
</tr>
<tr>
<td>Toutkoushian, Emily</td>
<td>1259</td>
</tr>
<tr>
<td>Trausan-Matu, Stefan</td>
<td>1663</td>
</tr>
<tr>
<td>Tscholl, Michael</td>
<td>1395</td>
</tr>
<tr>
<td>Tsunakawa, Takashi</td>
<td>1757</td>
</tr>
<tr>
<td>Tu, Xintian</td>
<td>1243, 1275, 1763</td>
</tr>
<tr>
<td>Tubman, Philip</td>
<td>1643</td>
</tr>
<tr>
<td>Tucker, Thomas</td>
<td>136</td>
</tr>
<tr>
<td>Turgeon, Krystle-Lee</td>
<td>1537</td>
</tr>
<tr>
<td>Tuvilla, Mavreen R.</td>
<td>1761</td>
</tr>
<tr>
<td>Tyler, Burr</td>
<td>1387</td>
</tr>
<tr>
<td>Tzou, Carrie</td>
<td>1187, 1753</td>
</tr>
<tr>
<td>Uttal, David</td>
<td>168, 961</td>
</tr>
<tr>
<td>Uttamchandani, Suraj</td>
<td>480, 1509</td>
</tr>
<tr>
<td>Vaiushampayan, Gauri</td>
<td>1495</td>
</tr>
<tr>
<td>Valcurel, Joshua</td>
<td>1387</td>
</tr>
<tr>
<td>Valcke, Martin</td>
<td>1296</td>
</tr>
<tr>
<td>van Aalst, Jan</td>
<td>176, 879, 1717, 1737, 1775</td>
</tr>
<tr>
<td>van Amersfoort, Daniël</td>
<td>1235</td>
</tr>
<tr>
<td>van Gog, Tamar</td>
<td>1681</td>
</tr>
<tr>
<td>van Heuven, Walter</td>
<td>969</td>
</tr>
<tr>
<td>Van Horne, Katie</td>
<td>656, 1211, 1513</td>
</tr>
<tr>
<td>Van Keer, Hilde</td>
<td>1296</td>
</tr>
<tr>
<td>van Leeuwen, Anouschka</td>
<td>1227</td>
</tr>
<tr>
<td>Vanlint, Alice</td>
<td>1153</td>
</tr>
<tr>
<td>Vanneste, Pieter</td>
<td>408</td>
</tr>
<tr>
<td>Varma, Keisha</td>
<td>913</td>
</tr>
<tr>
<td>Varma, Sashank</td>
<td>1073</td>
</tr>
<tr>
<td>Vea, Tanner</td>
<td>1187, 1659</td>
</tr>
<tr>
<td>Vedder-Weiss, Dana</td>
<td>41</td>
</tr>
<tr>
<td>Veldez, Rafael</td>
<td>64</td>
</tr>
<tr>
<td>Vempaty, Aditya</td>
<td>1391</td>
</tr>
<tr>
<td>Vens, Celine</td>
<td>408</td>
</tr>
<tr>
<td>Vieira, Camilo</td>
<td>1359</td>
</tr>
<tr>
<td>Villanosis, Krystal</td>
<td>1053</td>
</tr>
<tr>
<td>Villanucci Smothers, Cinzia</td>
<td>1415</td>
</tr>
<tr>
<td>Vincent, Marie-Caroline</td>
<td>1320</td>
</tr>
<tr>
<td>Virk, Satyugit</td>
<td>1259</td>
</tr>
<tr>
<td>Vitale, Jonathan</td>
<td>17, 1393</td>
</tr>
<tr>
<td>Vo, Tina</td>
<td>1397</td>
</tr>
<tr>
<td>Vogelstein, Lauren</td>
<td>1195, 1693</td>
</tr>
<tr>
<td>Vondracek, Mark</td>
<td>961</td>
</tr>
<tr>
<td>Vosniadou, Stella</td>
<td>1328</td>
</tr>
<tr>
<td>Vossoughi, Shirin</td>
<td>1283, 1312</td>
</tr>
<tr>
<td>Voyer, Samantha</td>
<td>1320</td>
</tr>
<tr>
<td>Wagh, Aditi</td>
<td>33, 1137</td>
</tr>
<tr>
<td>Wakefield, Wendy</td>
<td>887, 1725</td>
</tr>
<tr>
<td>Walker, Erin</td>
<td>737</td>
</tr>
<tr>
<td>Walker, Justice Toriba</td>
<td>1275, 1559</td>
</tr>
<tr>
<td>Walker, Susan</td>
<td>1589</td>
</tr>
<tr>
<td>Walkington, Candace A.</td>
<td>552, 1243, 1591</td>
</tr>
<tr>
<td>Wallace, Scott R.</td>
<td>1165</td>
</tr>
<tr>
<td>Wallon, Robert C.</td>
<td>1363, 1697</td>
</tr>
<tr>
<td>Walsh, Elizabeth</td>
<td>1493, 1727</td>
</tr>
<tr>
<td>Wan, Sally Wai Yan</td>
<td>1097</td>
</tr>
<tr>
<td>Wan, Sancia Wai-San</td>
<td>1097</td>
</tr>
<tr>
<td>Wang, Jing</td>
<td>176, 1351</td>
</tr>
<tr>
<td>Wang, Min</td>
<td>1591</td>
</tr>
<tr>
<td>Wang, Quichen</td>
<td>1371</td>
</tr>
<tr>
<td>Wang, Xu</td>
<td>296</td>
</tr>
<tr>
<td>Wang, Yu-Chi</td>
<td>913</td>
</tr>
<tr>
<td>Warchola, Gary</td>
<td>977</td>
</tr>
<tr>
<td>Warren, Beth</td>
<td>1283</td>
</tr>
<tr>
<td>Washington, Peter</td>
<td>1275</td>
</tr>
<tr>
<td>Weatherly, John</td>
<td>1437</td>
</tr>
<tr>
<td>Webb, Vanessa</td>
<td>1665</td>
</tr>
<tr>
<td>Weidler-Lewis, Joanna</td>
<td>1501</td>
</tr>
<tr>
<td>Weinberger, Armin</td>
<td>1296</td>
</tr>
<tr>
<td>Weintrup, David</td>
<td>1304</td>
</tr>
<tr>
<td>Wekerle, Christina</td>
<td>1569</td>
</tr>
<tr>
<td>Wen, Yun</td>
<td>207</td>
</tr>
<tr>
<td>Wendell, Kristen</td>
<td>1413</td>
</tr>
<tr>
<td>Wheeler, Penny</td>
<td>1663</td>
</tr>
<tr>
<td>White, Geneva</td>
<td>1671</td>
</tr>
<tr>
<td>White, Sarah</td>
<td>1065</td>
</tr>
<tr>
<td>Whittfield, Latricia</td>
<td>1235</td>
</tr>
<tr>
<td>Whitting, Justin</td>
<td>1519</td>
</tr>
<tr>
<td>Whittaker, Chris</td>
<td>1651</td>
</tr>
<tr>
<td>Wilde, Lorin</td>
<td>1391</td>
</tr>
<tr>
<td>Wilensky, Uri</td>
<td>871, 1304, 1617</td>
</tr>
<tr>
<td>Wilkerson, Michelle</td>
<td>1121, 1383</td>
</tr>
<tr>
<td>Williams, Christopher A.</td>
<td>1296</td>
</tr>
<tr>
<td>Williams, Joseph Jay</td>
<td>729, 1267</td>
</tr>
<tr>
<td>Wilson, Caitlin</td>
<td>1441</td>
</tr>
<tr>
<td>Wingert, Kerri</td>
<td>1211</td>
</tr>
<tr>
<td>Wink, Donald J.</td>
<td>1521, 1721</td>
</tr>
<tr>
<td>Wise, Alyssa</td>
<td>1491, 1799</td>
</tr>
<tr>
<td>Wisittanawat, Panchompoon</td>
<td>1335</td>
</tr>
<tr>
<td>Witerspoon, Eben</td>
<td>1443</td>
</tr>
<tr>
<td>Wong-Loke, Constance</td>
<td>96</td>
</tr>
<tr>
<td>Woods, Dawn</td>
<td>552, 1591</td>
</tr>
<tr>
<td>Woods, Philip</td>
<td>871</td>
</tr>
<tr>
<td>Worsley, Marcelo</td>
<td>1109, 1759</td>
</tr>
<tr>
<td>Wright, Casey E.</td>
<td>1761</td>
</tr>
<tr>
<td>Wu, Jue</td>
<td>961</td>
</tr>
<tr>
<td>Wu, Sally P.W.</td>
<td>504</td>
</tr>
<tr>
<td>Wylie, Alison</td>
<td>1251</td>
</tr>
<tr>
<td>Xia, Fangli</td>
<td>1483</td>
</tr>
<tr>
<td>Xia, Yu</td>
<td>1573</td>
</tr>
<tr>
<td>Xie, Bin</td>
<td>608</td>
</tr>
<tr>
<td>Xie, Charles</td>
<td>1359</td>
</tr>
</tbody>
</table>
Xie, Kui, 905, 965
Xing, Wanli, 855
Xu, Cindy, 895
Yamamoto, Ryota, 1507
Yamauchi, Yuhei, 1507
Yang, Chenglei, 1371
Yang, Hui, 368
Yang, Yuqin, 1717, 1737
Yao, Zheng, 1415
Yip, Jason, 705
Yoon, Susan, 224, 256, 1235
Young, Rose Nicole, 1665
Yu, Jinjian, 1769
Yuan, Guangji, 1033
Yung, Benny, 1565
Zahn, Miranda Renee, 264
Zaidi, Sania Zahra, 1211
Zambrano R., Jimmy, 112
Zaner, Sarah, 1687
Zeller, Laura, 1521
Zepeda, Cristina D., 1653
Zhang, Chao, 1651
Zhang, Heyi, 1101
Zhang, Jianwei, 484, 1033, 1133
Zhou, Chao, 1371
Zhou, He, 1033, 1133
Zhu, Gaoxia, 855, 1719
Zivic, Aliza, 25
Zohar, Asnat R., 1243, 1355, 1888
Keyword Index

Pages 1-670: Volume 1
Pages 671-1342: Volume 2
Pages 1343-1888: Volume 3

21st century skills................................................................................................................................. 448, 1629
3D display technologies.......................................................................................................................... 1371
3D learning........................................................................................................................................... 1495
academic achievement goals.................................................................................................................. 1351
accessibility .......................................................................................................................................... 1755
achievement disparities .......................................................................................................................... 1345
achievement emotions ......................................................................................................................... 1365
achievement goals................................................................................................................................. 965, 1561
action competence................................................................................................................................. 1441
action logs analysis ............................................................................................................................... 312
action research ...................................................................................................................................... 1525
active learning classrooms .................................................................................................................... 1651
activity systems analysis ....................................................................................................................... 1717
actor network mapping ......................................................................................................................... 656
Actor-Network Theory .......................................................................................................................... 1203
adaptive learning .................................................................................................................................. 729
ADDIE................................................................................................................................................... 1571
adolescent literacy ................................................................................................................................. 440, 1485
adolescents.......................................................................................................................................... 1449
adult learning ........................................................................................................................................ 1395
advanced learning technology ............................................................................................................. 1843
affect....................................................................................................................................................... 1105, 1585, 1843
affective states ...................................................................................................................................... 1691
affective technology .............................................................................................................................. 1843
affinity spaces ....................................................................................................................................... 824, 1523
affordances .......................................................................................................................................... 1275
African.................................................................................................................................................... 1611
after school contexts ............................................................................................................................. 48, 1513, 1761
agency .................................................................................................................................................... 1153, 1177, 1320, 1455, 1493, 1661
agent-based models .............................................................................................................................. 200, 949, 1057, 1617
algebra .................................................................................................................................................. 1705
alternative education ............................................................................................................................. 1613, 1619
ambient intelligence ............................................................................................................................ 1373
ambitious instruction ............................................................................................................................ 1451
analogy .................................................................................................................................................. 1065, 1607
anchoring ............................................................................................................................................... 1529
apprenticeship learning ......................................................................................................................... 1517
architectures for learning ..................................................................................................................... 96, 1533
area measurement ............................................................................................................................... 863
argumentation ....................................................................................................................................... 224, 400, 456, 576, 768, 977, 1479, 1601
artifact analysis ...................................................................................................................................... 1407
arts integration ....................................................................................................................................... 1621
assessment ........................................................................................................................................... 384, 576, 1304, 1369, 1381, 1439, 1521, 1537, 1713
at-risk .................................................................................................................................................... 1613
attraction/repulsion forces ................................................................................................................... 1355
augmented reality ................................................................................................................................. 136, 392
authentic learning ................................................................................................................................. 136, 1061
authentic science inquiry ....................................................................................................................... 997
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>autism</td>
<td>973</td>
</tr>
<tr>
<td>automated formative feedback</td>
<td>1577</td>
</tr>
<tr>
<td>axiological innovations</td>
<td>1187</td>
</tr>
<tr>
<td>badges</td>
<td>1753</td>
</tr>
<tr>
<td>bifocal modeling</td>
<td>1685</td>
</tr>
<tr>
<td>big data</td>
<td>1791</td>
</tr>
<tr>
<td>biology</td>
<td>1275</td>
</tr>
<tr>
<td>black girls</td>
<td>1181</td>
</tr>
<tr>
<td>blended learning</td>
<td>632, 1545</td>
</tr>
<tr>
<td>block-based programming</td>
<td>328</td>
</tr>
<tr>
<td>board games</td>
<td>913, 1611, 1743</td>
</tr>
<tr>
<td>boundary crossing</td>
<td>981, 1001</td>
</tr>
<tr>
<td>broadening participation</td>
<td>328</td>
</tr>
<tr>
<td>brokering</td>
<td>1671</td>
</tr>
<tr>
<td>card sort</td>
<td>1505</td>
</tr>
<tr>
<td>careers</td>
<td>1475, 1605</td>
</tr>
<tr>
<td>case study</td>
<td>440, 957, 1335, 1447, 1559, 1603, 1605, 1693</td>
</tr>
<tr>
<td>CATLM</td>
<td>1411</td>
</tr>
<tr>
<td>challenge-based learning; situated cognition; social network analysis; CSCL</td>
<td>1069</td>
</tr>
<tr>
<td>chemical bonding</td>
<td>1312</td>
</tr>
<tr>
<td>chemistry</td>
<td>1355</td>
</tr>
<tr>
<td>chemistry labs</td>
<td>504</td>
</tr>
<tr>
<td>child-led learning</td>
<td>1711</td>
</tr>
<tr>
<td>Chinese character learning</td>
<td>208</td>
</tr>
<tr>
<td>chronotopic analysis</td>
<td>1041</td>
</tr>
<tr>
<td>citizen science</td>
<td>1499, 1517, 1641, 1645</td>
</tr>
<tr>
<td>civic education</td>
<td>1593, 1695</td>
</tr>
<tr>
<td>classification</td>
<td>1491</td>
</tr>
<tr>
<td>classroom</td>
<td>1559</td>
</tr>
<tr>
<td>classroom assessment</td>
<td>528, 1211</td>
</tr>
<tr>
<td>classroom culture</td>
<td>1447</td>
</tr>
<tr>
<td>classroom discourse</td>
<td>232, 1477, 1769</td>
</tr>
<tr>
<td>classroom discussion</td>
<td>440, 1485</td>
</tr>
<tr>
<td>classroom dynamics</td>
<td>1769</td>
</tr>
<tr>
<td>classroom efficacy study</td>
<td>1647</td>
</tr>
<tr>
<td>classroom orchestration</td>
<td>104</td>
</tr>
<tr>
<td>classroom video</td>
<td>1005</td>
</tr>
<tr>
<td>climate change</td>
<td>1493, 1715</td>
</tr>
<tr>
<td>co-design</td>
<td>248, 520, 1025, 1157, 1749</td>
</tr>
<tr>
<td>coding</td>
<td>1571, 1739</td>
</tr>
<tr>
<td>cognition</td>
<td>17, 336, 464, 1393, 1581, 1585</td>
</tr>
<tr>
<td>cognitive apprenticeship</td>
<td>1573, 1741</td>
</tr>
<tr>
<td>cognitive challenges</td>
<td>1133</td>
</tr>
<tr>
<td>cognitive diagnosis</td>
<td>1435</td>
</tr>
<tr>
<td>cognitive load theory</td>
<td>112, 1423, 1567</td>
</tr>
<tr>
<td>cognitive-situative debate</td>
<td>1328</td>
</tr>
<tr>
<td>coherence</td>
<td>25, 1729</td>
</tr>
<tr>
<td>collaboration</td>
<td>272, 448, 536, 1109, 1379, 1459, 1461, 1503, 1521, 1549, 1629, 1631</td>
</tr>
<tr>
<td>collaboration scripts</td>
<td>504</td>
</tr>
<tr>
<td>collaboration-as-learning</td>
<td>48</td>
</tr>
<tr>
<td>collaborative dialogs</td>
<td>1445</td>
</tr>
<tr>
<td>collaborative discussion</td>
<td>977</td>
</tr>
<tr>
<td>collaborative inquiry</td>
<td>855, 1729</td>
</tr>
<tr>
<td>collaborative knowledge construction</td>
<td>496</td>
</tr>
<tr>
<td>collaborative learning</td>
<td>112, 208, 312, 608, 745, 887, 921, 1045, 1149, 1535, 1541, 1561, 1663, 1725, 1733</td>
</tr>
<tr>
<td>collaborative problem solving</td>
<td>144, 496, 1557</td>
</tr>
<tr>
<td>Concept</td>
<td>Page Numbers</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Collaborative rationality</td>
<td>945</td>
</tr>
<tr>
<td>Collective knowledge advancement</td>
<td>1757, 1775</td>
</tr>
<tr>
<td>Communication</td>
<td>1367</td>
</tr>
<tr>
<td>Communication skills</td>
<td>721</td>
</tr>
<tr>
<td>Communities of practice</td>
<td>985, 1415, 1589, 1597</td>
</tr>
<tr>
<td>Community ethnography</td>
<td>80</td>
</tr>
<tr>
<td>Community science</td>
<td>1499, 1517</td>
</tr>
<tr>
<td>Community-based collaborative design partnerships</td>
<td>1312</td>
</tr>
<tr>
<td>Community-based learning</td>
<td>705</td>
</tr>
<tr>
<td>Competency-based education</td>
<td>792</td>
</tr>
<tr>
<td>Complex systems</td>
<td>656, 1395</td>
</tr>
<tr>
<td>Complex task</td>
<td>933</td>
</tr>
<tr>
<td>Complexity</td>
<td>1328</td>
</tr>
<tr>
<td>Complex-systems</td>
<td>949, 1057</td>
</tr>
<tr>
<td>Comprehension</td>
<td>152</td>
</tr>
<tr>
<td>Computational action</td>
<td>1657</td>
</tr>
<tr>
<td>Computational discourse</td>
<td>1745</td>
</tr>
<tr>
<td>Computational literacy</td>
<td>544</td>
</tr>
<tr>
<td>Computational modeling</td>
<td>33, 1137</td>
</tr>
<tr>
<td>Computational thinking</td>
<td>368, 384, 544, 721, 871, 1121, 1304, 1439, 1443, 1503, 1579, 1657, 1739</td>
</tr>
<tr>
<td>Computer games</td>
<td>1507</td>
</tr>
<tr>
<td>Computer science education</td>
<td>328, 448, 544, 640, 925, 1085, 1377, 1429, 1447, 1531, 1557</td>
</tr>
<tr>
<td>Computer simulations</td>
<td>.56</td>
</tr>
<tr>
<td>Computer-mediated communication</td>
<td>1375, 1491</td>
</tr>
<tr>
<td>Computer-supported collaborative problem-solving</td>
<td>184</td>
</tr>
<tr>
<td>Computing education</td>
<td>560, 1304, 1657</td>
</tr>
<tr>
<td>Concept learning</td>
<td>1117</td>
</tr>
<tr>
<td>Conceptual blending</td>
<td>957</td>
</tr>
<tr>
<td>Conceptual change</td>
<td>400, 729</td>
</tr>
<tr>
<td>Conceptual design</td>
<td>1399</td>
</tr>
<tr>
<td>Conceptual goals</td>
<td>1697</td>
</tr>
<tr>
<td>Conceptual knowledge</td>
<td>933</td>
</tr>
<tr>
<td>Conceptual understanding</td>
<td>176</td>
</tr>
<tr>
<td>Connected learning</td>
<td>705, 1605</td>
</tr>
<tr>
<td>Connective learning</td>
<td>1145</td>
</tr>
<tr>
<td>Consequential engagement</td>
<td>887</td>
</tr>
<tr>
<td>Constructionism</td>
<td>488, 921, 1165, 1575, 1617, 1743</td>
</tr>
<tr>
<td>Constructivism</td>
<td>72, 792</td>
</tr>
<tr>
<td>Content creation</td>
<td>1267</td>
</tr>
<tr>
<td>Content validity</td>
<td>1505</td>
</tr>
<tr>
<td>Control-of-variable-strategy</td>
<td>1601</td>
</tr>
<tr>
<td>Conversation analysis</td>
<td>1085, 1405</td>
</tr>
<tr>
<td>Conversational learning</td>
<td>1391</td>
</tr>
<tr>
<td>Cooperative learning</td>
<td>913</td>
</tr>
<tr>
<td>Co-regulation</td>
<td>1296, 1347</td>
</tr>
<tr>
<td>Craft education</td>
<td>1417</td>
</tr>
<tr>
<td>Creative computing</td>
<td>592</td>
</tr>
<tr>
<td>Creativity</td>
<td>1419</td>
</tr>
<tr>
<td>Critical</td>
<td>1105</td>
</tr>
<tr>
<td>Critical historical perspective</td>
<td>192</td>
</tr>
<tr>
<td>Critical theory</td>
<td>1515</td>
</tr>
<tr>
<td>Critical thinking</td>
<td>1473</td>
</tr>
<tr>
<td>Critique</td>
<td>816, 1049</td>
</tr>
<tr>
<td>Cross-curricular</td>
<td>1719</td>
</tr>
<tr>
<td>Crosscutting concepts</td>
<td>1719</td>
</tr>
<tr>
<td>Cross-disciplinary collaboration</td>
<td>1707</td>
</tr>
<tr>
<td>Cross-setting learning</td>
<td>1671</td>
</tr>
<tr>
<td>Term</td>
<td>Pages</td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>crowdsourcing</td>
<td>1267</td>
</tr>
<tr>
<td>CSCL</td>
<td>88, 296, 336, 496, 855, 905, 965, 1045, 1161, 1173, 1259, 1389, 1457, 1489, 1491, 1507, 1555, 1593, 1657, 1695, 1737, 1755, 1831</td>
</tr>
<tr>
<td>CSCL discourse</td>
<td>879, 1775</td>
</tr>
<tr>
<td>cultural-historical activity theory</td>
<td>1113, 1153, 1621</td>
</tr>
<tr>
<td>culturally responsive</td>
<td>697</td>
</tr>
<tr>
<td>culture</td>
<td>1665</td>
</tr>
<tr>
<td>culture and learning</td>
<td>1627</td>
</tr>
<tr>
<td>CURE</td>
<td>1361</td>
</tr>
<tr>
<td>curriculum</td>
<td>1017</td>
</tr>
<tr>
<td>dashboards</td>
<td>104</td>
</tr>
<tr>
<td>data analysis</td>
<td>1383</td>
</tr>
<tr>
<td>data literacy</td>
<td>1113</td>
</tr>
<tr>
<td>data mining</td>
<td>855</td>
</tr>
<tr>
<td>data representation</td>
<td>1001</td>
</tr>
<tr>
<td>data science education</td>
<td>1383</td>
</tr>
<tr>
<td>data structures</td>
<td>1539</td>
</tr>
<tr>
<td>data visualization</td>
<td>144</td>
</tr>
<tr>
<td>database search</td>
<td>1551</td>
</tr>
<tr>
<td>deaf/hard-of-hearing</td>
<td>1623</td>
</tr>
<tr>
<td>debugging</td>
<td>1405</td>
</tr>
<tr>
<td>deeper learning</td>
<td>925</td>
</tr>
<tr>
<td>design</td>
<td>1259, 1419, 1617, 1655</td>
</tr>
<tr>
<td>design learning</td>
<td>352, 1379, 1663</td>
</tr>
<tr>
<td>design pattern</td>
<td>1545</td>
</tr>
<tr>
<td>design principles</td>
<td>1519</td>
</tr>
<tr>
<td>design research</td>
<td>1177, 1391, 1419</td>
</tr>
<tr>
<td>design thinking</td>
<td>1529, 1663</td>
</tr>
<tr>
<td>design-based implementation research</td>
<td>96, 673, 1157, 1525, 1687, 1729, 1807</td>
</tr>
<tr>
<td>design-based innovation networks</td>
<td>96</td>
</tr>
<tr>
<td>design-based research</td>
<td>9, 288, 320, 432, 488, 528, 584, 648, 808, 824, 863, 887, 917, 925, 961, 1009, 1025, 1137, 1243, 1275, 1361, 1397, 1403, 1413, 1415, 1425, 1427, 1433, 1459, 1487, 1489, 1513, 1529, 1543, 1579, 1637, 1643, 1709, 1773, 1779</td>
</tr>
<tr>
<td>designed learning environments</td>
<td>616</td>
</tr>
<tr>
<td>development</td>
<td>1581</td>
</tr>
<tr>
<td>dialectical agency</td>
<td>1711</td>
</tr>
<tr>
<td>dialogic instruction</td>
<td>1709</td>
</tr>
<tr>
<td>dialogic literacy</td>
<td>945</td>
</tr>
<tr>
<td>dialogical theory</td>
<td>1535</td>
</tr>
<tr>
<td>dialogue</td>
<td>737</td>
</tr>
<tr>
<td>dialogue videos</td>
<td>953</td>
</tr>
<tr>
<td>differential sequence mining</td>
<td>1691</td>
</tr>
<tr>
<td>digital badges</td>
<td>792</td>
</tr>
<tr>
<td>digital identity</td>
<td>1731</td>
</tr>
<tr>
<td>digital interaction</td>
<td>1463</td>
</tr>
<tr>
<td>digital journalism</td>
<td>1049</td>
</tr>
<tr>
<td>digital learning environments</td>
<td>1551, 1771, 1773</td>
</tr>
<tr>
<td>digital literacies</td>
<td>1129, 1449</td>
</tr>
<tr>
<td>digital reading</td>
<td>1435</td>
</tr>
<tr>
<td>digital story telling</td>
<td>1403</td>
</tr>
<tr>
<td>digital text</td>
<td>240</td>
</tr>
<tr>
<td>digital tools</td>
<td>1641</td>
</tr>
<tr>
<td>dignity</td>
<td>480</td>
</tr>
<tr>
<td>DNA</td>
<td>1435</td>
</tr>
<tr>
<td>disciplinary practices</td>
<td>1290, 1649</td>
</tr>
<tr>
<td>discourse</td>
<td>847, 895, 1053, 1509, 1721, 1735</td>
</tr>
<tr>
<td>discourse analysis</td>
<td>600, 760, 1401, 1477, 1619, 1625, 1771</td>
</tr>
</tbody>
</table>
epistemic criteria .................................................................................................................. 663
epistemic engagement ........................................................................................................... 1479
epistemic heterogeneity ....................................................................................................... 1283
episodic network analysis ..................................................................................................... 144, 871, 1353, 1625
episodic objects .................................................................................................................... 248, 784
episodic practices .................................................................................................................. 288
episodic reflection .................................................................................................................. 240
episodic understanding ......................................................................................................... 176
episodic virtues .................................................................................................................... 1467
epistemological framing ................................................................................................. 1121, 1137
epistemology ....................................................................................................................... 997, 1367
equitable design .................................................................................................................... 1187
equity ................................................................................................................................. 80, 88, 480, 488, 656, 909, 1021, 1053, 1081, 1141, 1149, 1429, 1515, 1671, 1693, 1729
equity ................................................................................................................................... 1527
ethic of caring ..................................................................................................................... 424
ethics and politics of learning ........................................................................................... 1283
ethnography ....................................................................................................................... 752, 1613, 1707
evaluate evidence ............................................................................................................... 176
everyday experiences ......................................................................................................... 989
everyday science ................................................................................................................ 1699
evidence ............................................................................................................................. 224, 1251
evidence centered design ................................................................................................. 384
evidence construction ........................................................................................................ 1383
evidence-based practice .................................................................................................... 1037
evidence-based reasoning ............................................................................................. 1037
example-based Learning ................................................................................................. 1681
expansive learning ........................................................................................................... 1312
expectancy value theory ................................................................................................. 472
experiential learning ........................................................................................................ 1371
experimental design ........................................................................................................ 800
expertise ............................................................................................................................ 1569, 1641
explanation ....................................................................................................................... 1363
explanatory power ............................................................................................................ 1117
exploratory question posing ............................................................................................ 1539
eye movements ................................................................................................................ 969
eye tracking ......................................................................................................................... 640, 1537
facilitation ......................................................................................................................... 592, 1759
failure ............................................................................................................................... 981
families ................................................................................................................................ 41, 592, 752
family engagement .......................................................................................................... 1409
family environment .......................................................................................................... 1767
family learning .................................................................................................................. 1187, 1753
feedback .......................................................................................................................... 1049, 1537, 1779
feedback criteria .............................................................................................................. 1577
flexible thinking ................................................................................................................ 1381
flow experience ................................................................................................................. 184
formative assessment .................................................................................................... 520, 1461, 1597, 1705, 1735
formative feedback ......................................................................................................... 1471
formative intervention research ..................................................................................... 1471
framing ............................................................................................................................. 1169
function-structure-behaviour .......................................................................................... 1455
future-scaffolding skills ................................................................................................. 1399
game based learning ....................................................................................................... 1441
game design ...................................................................................................................... 1349, 1575, 1743
game mechanics ............................................................................................................. 1385
game-based formative feedback .................................................................................... 1471
game-based language learning ................................................................. 208
game-based learning .............................................................................. 648, 1487
gameplay .................................................................................................. 1747
games ........................................................................................................ 913, 1523, 1745
gamification ............................................................................................. 1471
gender ......................................................................................................... 800, 1509, 1835
genetics ....................................................................................................... 576, 808
geographies of responsibility ................................................................. 1283
geology ......................................................................................................... 1599
geometry .................................................................................................. 1165
gesture ......................................................................................................... 552, 929, 1085, 1379, 1591
gesture-augmented computer simulations ............................................... 1363
girls ............................................................................................................. 1503
girls-only ................................................................................................... 1661
goals ........................................................................................................... 1469
graduate attributes .................................................................................... 941
grounded theory ......................................................................................... 1033
group awareness ........................................................................................ 456, 1457, 1555
group communication analysis ............................................................... 1815
group formation ......................................................................................... 1555
group interaction ....................................................................................... 1815
growth mindset .......................................................................................... 1033
guided participation .................................................................................. 1093
health education ......................................................................................... 949
help seeking ............................................................................................... 472
heterogeneity ............................................................................................. 480
higher education ........................................................................................ 536, 1427, 1551, 1577, 1713
historical inquiry ......................................................................................... 136
historical thinking .................................................................................... 1483
holistic ....................................................................................................... 376
home environments ................................................................................... 752
human centered robotics ......................................................................... 1407
human computer interaction ..................................................................... 88
human robot interaction ........................................................................... 88, 1675
human-centered robotics ......................................................................... 1563, 1609
humanities ............................................................................................... 945, 1535
human-technology .................................................................................. 1645
idea analysis .............................................................................................. 1401
idea diversity ............................................................................................ 1013
idea emergence ........................................................................................ 1757
idea improvement ..................................................................................... 1013
Identity .................................................................................................... 41, 120, 800, 1427, 1493, 1499, 1613, 1619, 1701, 1721, 1751, 1753, 1773, 1753
identity development .............................................................................. 1773
identity resources ..................................................................................... 1753
ideology .................................................................................................... 1219, 1655
immersion ................................................................................................. 392
immersive environment .......................................................................... 941
impact ....................................................................................................... 568
implementation ........................................................................................ 1387
improvisation .......................................................................................... 72
incomplete example ................................................................................ 1529
Indigenous ............................................................................................... 697
individual differences .............................................................................. 1389
indoor positioning ................................................................................... 1077
infographics ............................................................................................. 1001
informal environments ........................................................................... 41
informal learning ..................................................................................... 592, 847, 1335, 1367, 1429, 1473, 1475, 1497, 1499, 1517, 1605, 1715
online communities ................................................................. 1523, 1589
online learning .................................................................................. 536, 965, 1009, 1029, 1375, 1395, 1511
online/distance education ............................................................. 1741
open education .................................................................................. 1029
open learning ...................................................................................... 536
open-ended learning environments .................................................. 304, 1691
open-mindedness .............................................................................. 1467
orchestration ..................................................................................... 1541, 1651
orchestration tools ............................................................................. 1227
ordering effect .................................................................................... 408
organizational factors ......................................................................... 1345
organizational learning ....................................................................... 1203
organizing social futures .................................................................... 1673
OST ................................................................................................... 1475
other-regulation .................................................................................. 1347
out-of-school learning ......................................................................... 624, 1061
outreach .............................................................................................. 120
pair programming ............................................................................. 1557
parenting ............................................................................................ 1589
parent-oriented goals ........................................................................... 1351
participation ........................................................................................ 1153
participatory design ........................................................................... 1161, 1541
participatory simulation ...................................................................... 512, 584, 1077
participatory-design research ............................................................ 1150
PBL .................................................................................................. 1483, 1831
post-lesson colloquium ........................................................................ 673
possible selves ..................................................................................... 1571
power and identity ................................................................................ 832
positioning .......................................................................................... 832, 1053, 1651
positioning theory .............................................................................. 1081, 1623
possible selves ..................................................................................... 1571
post-lesson colloquium ........................................................................ 673
power and identity ................................................................................ 832
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>self regulated learning</td>
<td>1691</td>
</tr>
<tr>
<td>self-directed learning</td>
<td></td>
</tr>
<tr>
<td>self-efficacy</td>
<td>1533</td>
</tr>
<tr>
<td>self-explanation</td>
<td>600, 737, 1365, 1593</td>
</tr>
<tr>
<td>self-explanation prompts</td>
<td>729, 1267</td>
</tr>
<tr>
<td>self-regulated learning</td>
<td>933</td>
</tr>
<tr>
<td>semiformal learning</td>
<td>1351, 1561</td>
</tr>
<tr>
<td>sensemaking</td>
<td>1041</td>
</tr>
<tr>
<td>sensory augmentation</td>
<td>1483</td>
</tr>
<tr>
<td>shared goal</td>
<td>1557</td>
</tr>
<tr>
<td>simulations</td>
<td>997, 1697</td>
</tr>
<tr>
<td>situated cognition</td>
<td>1699</td>
</tr>
<tr>
<td>situated learning</td>
<td>1659, 1763</td>
</tr>
<tr>
<td>situational interest</td>
<td>1061</td>
</tr>
<tr>
<td>sketching</td>
<td>937</td>
</tr>
<tr>
<td>small group</td>
<td>1737</td>
</tr>
<tr>
<td>smart classrooms</td>
<td>1779</td>
</tr>
<tr>
<td>smartwatch</td>
<td>1699</td>
</tr>
<tr>
<td>social and emotional learning</td>
<td>913</td>
</tr>
<tr>
<td>social capital</td>
<td>1235</td>
</tr>
<tr>
<td>social comparison</td>
<td>1421</td>
</tr>
<tr>
<td>social experience</td>
<td>737</td>
</tr>
<tr>
<td>social justice</td>
<td>80, 713</td>
</tr>
<tr>
<td>social learning</td>
<td>1589, 1643</td>
</tr>
<tr>
<td>social media</td>
<td>344, 416, 456, 1511</td>
</tr>
<tr>
<td>social metacognition</td>
<td>1633</td>
</tr>
<tr>
<td>social negotiation</td>
<td>1479</td>
</tr>
<tr>
<td>social network analysis</td>
<td>905, 1347, 1375, 1403, 1711, 1725, 1775</td>
</tr>
<tr>
<td>social networks</td>
<td>1235</td>
</tr>
<tr>
<td>social regulation of learning</td>
<td>1631</td>
</tr>
<tr>
<td>Social semiotics</td>
<td>1320</td>
</tr>
<tr>
<td>social skills training</td>
<td>973</td>
</tr>
<tr>
<td>social status goals</td>
<td>1351</td>
</tr>
<tr>
<td>socialization</td>
<td>1741</td>
</tr>
<tr>
<td>socially acute questions</td>
<td>456</td>
</tr>
<tr>
<td>socially shared regulation</td>
<td>1561</td>
</tr>
<tr>
<td>sociocognitive roles</td>
<td>1815</td>
</tr>
<tr>
<td>sociocultural learning</td>
<td>776</td>
</tr>
<tr>
<td>sociocultural perspective</td>
<td>989</td>
</tr>
<tr>
<td>sociocultural theory</td>
<td>1320, 1515</td>
</tr>
<tr>
<td>socio-emotional learning</td>
<td>1573</td>
</tr>
<tr>
<td>socio-emotional regulation</td>
<td>1013</td>
</tr>
<tr>
<td>socio-emotional responses</td>
<td>1133</td>
</tr>
<tr>
<td>sociomaterial</td>
<td>1643</td>
</tr>
<tr>
<td>socio-political dimensions of learning</td>
<td>1673</td>
</tr>
<tr>
<td>socioscientific issues</td>
<td>256, 1715</td>
</tr>
<tr>
<td>socio-semantic network analysis</td>
<td>673, 1757</td>
</tr>
<tr>
<td>software</td>
<td>1399</td>
</tr>
<tr>
<td>Songo</td>
<td>1611</td>
</tr>
<tr>
<td>spatial thinking</td>
<td>168</td>
</tr>
<tr>
<td>statistics</td>
<td>1121, 1365</td>
</tr>
<tr>
<td>statistics instruction</td>
<td>1507</td>
</tr>
<tr>
<td>STEAM</td>
<td>168, 1203, 1707</td>
</tr>
<tr>
<td>STEAM education</td>
<td>1549</td>
</tr>
<tr>
<td>STEM</td>
<td>64, 120, 855, 909, 993, 1041, 1105, 1187, 1259, 1469, 1501, 1665, 1701</td>
</tr>
<tr>
<td>STEM education</td>
<td>1437, 1441</td>
</tr>
<tr>
<td>STEM learning</td>
<td>608, 1407, 1563</td>
</tr>
</tbody>
</table>
### Index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEM teaching</td>
<td>1113</td>
</tr>
<tr>
<td>step duration</td>
<td>993</td>
</tr>
<tr>
<td>student affect</td>
<td>25, 1385</td>
</tr>
<tr>
<td>student engagement</td>
<td>1447</td>
</tr>
<tr>
<td>student inquiry</td>
<td>1145</td>
</tr>
<tr>
<td>student learning</td>
<td>863</td>
</tr>
<tr>
<td>student modeling</td>
<td>993</td>
</tr>
<tr>
<td>student perceptions</td>
<td>1683</td>
</tr>
<tr>
<td>student reasoning with relationally-specified quantities</td>
<td>1637</td>
</tr>
<tr>
<td>student voice</td>
<td>1161</td>
</tr>
<tr>
<td>student-centered teaching</td>
<td>160</td>
</tr>
<tr>
<td>study strategies</td>
<td>1653</td>
</tr>
<tr>
<td>subject-matter learning</td>
<td>1757</td>
</tr>
<tr>
<td>survey</td>
<td>472</td>
</tr>
<tr>
<td>synthetic biology</td>
<td>1559</td>
</tr>
<tr>
<td>systematic review</td>
<td>1097</td>
</tr>
<tr>
<td>systems thinking</td>
<td>512, 584, 917, 1077</td>
</tr>
<tr>
<td>tabletop</td>
<td>1747</td>
</tr>
<tr>
<td>talk moves</td>
<td>1709</td>
</tr>
<tr>
<td>tangible model</td>
<td>1685</td>
</tr>
<tr>
<td>tangible user interfaces</td>
<td>208, 1459, 1531</td>
</tr>
<tr>
<td>taxonomy of learning</td>
<td>985</td>
</tr>
<tr>
<td>teachable robot</td>
<td>737</td>
</tr>
<tr>
<td>teacher attitudes</td>
<td>1101</td>
</tr>
<tr>
<td>teacher awareness tools</td>
<td>104</td>
</tr>
<tr>
<td>teacher beliefs and attitudes</td>
<td>160</td>
</tr>
<tr>
<td>teacher co-design</td>
<td>663</td>
</tr>
<tr>
<td>teacher development</td>
<td>1731</td>
</tr>
<tr>
<td>teacher education</td>
<td>368, 648, 713, 1037, 1235, 1353, 1357, 1403, 1429, 1515, 1597, 1633</td>
</tr>
<tr>
<td>teacher improvisation</td>
<td>232</td>
</tr>
<tr>
<td>teacher knowledge</td>
<td>368, 1481</td>
</tr>
<tr>
<td>teacher learning</td>
<td>713, 760, 1009, 1177, 1511, 1689</td>
</tr>
<tr>
<td>teacher mediation</td>
<td>280</td>
</tr>
<tr>
<td>teacher noticing and responsiveness</td>
<td>1017</td>
</tr>
<tr>
<td>teacher orchestration</td>
<td>1227</td>
</tr>
<tr>
<td>teacher presence</td>
<td>1631</td>
</tr>
<tr>
<td>teacher professional development</td>
<td>1157, 1677, 1769</td>
</tr>
<tr>
<td>teacher professional learning</td>
<td>824, 1005</td>
</tr>
<tr>
<td>teacher reform capacity</td>
<td>1533</td>
</tr>
<tr>
<td>teacher support</td>
<td>1541</td>
</tr>
<tr>
<td>teachers as designers</td>
<td>1749</td>
</tr>
<tr>
<td>teachers change</td>
<td>1749</td>
</tr>
<tr>
<td>teaching</td>
<td>1443, 1495, 1579, 1651, 1653</td>
</tr>
<tr>
<td>teaching and learning processes</td>
<td>408</td>
</tr>
<tr>
<td>teaching assistant</td>
<td>1451</td>
</tr>
<tr>
<td>team leadership</td>
<td>905, 965</td>
</tr>
<tr>
<td>team performance</td>
<td>144</td>
</tr>
<tr>
<td>teamwork</td>
<td>941, 1461</td>
</tr>
<tr>
<td>technological context</td>
<td>1491</td>
</tr>
<tr>
<td>technology</td>
<td>632, 752, 816, 863, 1705, 1723, 1735, 1751</td>
</tr>
<tr>
<td>technology and education</td>
<td>17, 1393</td>
</tr>
<tr>
<td>technology and research</td>
<td>1391</td>
</tr>
<tr>
<td>technology enhanced assessment</td>
<td>1783</td>
</tr>
<tr>
<td>technology integration</td>
<td>1731</td>
</tr>
<tr>
<td>technology use</td>
<td>160</td>
</tr>
<tr>
<td>technology-based teaching</td>
<td>1553</td>
</tr>
<tr>
<td>technology-enhanced classroom</td>
<td>1437</td>
</tr>
</tbody>
</table>