Cross-Cultural Generalizability of Social and Dimensional Comparison Effects on Reading, Math, and Science Self-Concepts for Primary School Students Using the Combined PIRLS and TIMSS Data

Jiesi Guo [a], Herbert W. Marsh [a,b], Philip D. Parker [a], Theresa Dicke[a]

[a] Australian Catholic University (Australia)

[b] Oxford University (UK)

Citation: Guo, J., Marsh, H. W., Parker, P. D., & Dicke, T. (2018). Cross-cultural generalizability of social and dimensional comparison effects on reading, math, and science self-concepts for primary school students using the combined PIRLS and TIMSS data. *Learning and Instruction*, *58*(July), 210–219.
https://doi.org/10.1016/j.learninstruc.2018.07.007

Author Note. This research was funded in part by grants from the Australian Research Council awarded to Herbert W. Marsh, Alexandre J. S. Morin and Philip D. Parker (DP130102713).

Correspondence concerning this article should be addressed to Jiesi Guo, Institute for Positive Psychology and Education, Australian Catholic University, 33 Berry Street, North Sydney NSW 2060 Email: jiesiguo@gmail.com

Abstract

Previous cross-cultural studies of social and dimensional comparison processes forming academic self-concepts (the big-fish-little-pond effect (BFLPE) and Internal-external frame-of-reference (I/E) models) have mostly been based on high-school students and two subject domains. Our study is the first to test the cross-cultural generalizability of both comparison processes across reading, mathematics, and science by combining of the TIMSS and PIRLS 2011 databases (15 OECD countries, 67,386 fourth-graders). Consistent with the I/E model, high achievement in mathematics/reading had positive effects on self-concept in the matching domain but negative effects in the non-matching domain. Extending the I/E model, students engaged in assimilating comparisons between science and reading (i.e., achievement in one subject had positive effects on self-concept in the other) but contrasting comparisons between mathematics and science. Strong BFLPEs (negative effects of class-average achievement on self-concept) were found for mathematics but were smaller for reading and science. The results generalized well across all countries.

Keywords: self-concept, big-fish-little-pond effect, Internal/external frame of reference model, dimensional comparisons, cross-cultural

Academic self-concept (ASC), defined as self-perceptions of one's academic abilities, is a key construct in developmental and educational psychology. ASC is positively linked to better knowledge acquisition, greater perseverance, and higher educational and occupational aspirations and attainment (Guo, Marsh, Morin, Parker, & Kaur, 2015; Guo, Parker, Marsh, & Morin, 2015, Guo et al., 2016; Marsh et al., 2017). The formation of ASC is highly responsive to dimensional and social comparison processes posited in two theoretical models: the internal/external frame of reference model (I/E) model and the big-fish-little-pond effect (BFLPE; Marsh, 2007). In the last two decades, an increasing number of cross-cultural studies have provided strong support for the generalizability of these two ASC models, using the Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA) data (e.g., Chiu, 2012; Guo, Marsh, Parker, Morin, & Dicke, 2017; Marsh & Hau, 2003, 2004; Seaton, Marsh, & Craven, 2009; Nagengast & Marsh, 2012). However, those cross-cultural studies are largely based on secondary school students (but see Marsh, Abduljabbar et al. 2015a, 2015b). Yet, primary school might be the most critical period in the formation of children's ASCs, given the rapid development of cognitive and academic skills (Harter, 2012). In particular, the cross-cultural generalizability of the I/E model in the verbal domain (e.g., reading) and mathematics, both basic building blocks of all academic subjects, have not been examined in primary-age children. Furthermore, little research has investigated the cross-cultural generalizability of the BFLPE in reading and science particularly for primary-age children. The main reason of these limitations is the lack of available cross-cultural data: PISA only focuses on students15 years of age and generally includes ASC in a single domain in each cycle; TIMSS draws on fourth- and eighth-graders but only assesses ASCs in mathematics and science.

In 2011, the five-year cycle of the Progress in International Reading Literacy Study (PIRLS) — another large-scale international assessment focussing on fourth-graders' reading competencies — came into alignment with the four-year cycle of TIMSS for the first time. This provides an unprecedented opportunity for researchers to merge these two databases and compare fourth-graders in three fundamental curricular areas: mathematics, science, and reading[1], across countries. Our study is apparently the first to take advantage of the combination of the TIMSS and PIRLS data to fill the critical gaps in research on the generalizability of strong ASC theoretical models (i.e., the I/E and BFLPE) in primary-age cohorts. In particular, we incorporate the two models within a unified theoretical framework and provide even stronger tests of the universality of ASC theories across reading, mathematics, and science. Therefore, the present study makes a unique contribution to the ASC literature in following ways: this study is the first to look at cross-cultural generalizability of (1) the I/E model in mathematics and reading for primary school students, (2) the DCT (Dimensional comparison theory) expansion of the I/E model across three core academic domains; and (3) the BFLPE in reading and science for primary school students.

Dimensional and Social Comparisons and Self-Concept Development

Existing research has indicated that children's ASCs decline in each subject domain following primary school and become relatively stable during late adolescence (see Wigfield, Tonks, & Klauda, 2016 for a review). Many researchers have attributed the declining ASC trajectory primarily to aspects of cognitive development and school environments (e.g., Eccles et al., 1993; Harter, 2012). Children in the early elementary years tend to be quite optimistic about their abilities in different subject domains and these beliefs are not substantially related to external indicators, such as skills and accomplishments (Marsh et al., 2015a ,2015b). However, with cognitive development children become more capable of

[1] In relation to reading curriculum, it is usually incorporated in the national language curriculum except for France, Hungary, the Netherlands which have a national curriculum specifically for reading (see Mullis *et al., 2012, PIRLS 2011 Encyclopaedia*). Although reading instructions vary substantially across countries, depending on resources, culture, and educational philosophies, reading is taught as part of the national language curriculum that also includes writing and other communication skills in all participating countries included in the current study.

interpreting and integrating ability-related experience such as feedback (e.g., grades and test scores) into ASC and of learning their relative strengths and weaknesses. Such self-evaluations rely on two comparison processes: social comparisons (by likening their performance in a subject domain with that of their peers in the same class/school) and dimensional comparisons (by contrasting their achievement in one domain with that in other domains (Marsh, 2007; Marsh et al., 2017). Dimensional and social comparison processes became more salient with age, and thus ASC across different subject domains becomes more differentiated and more influenced by school feedback and academic environment (Dicke et al., 2018; Ehm, Lindberg, & Hasselhorn, 2014; Schmidt, & Brunner, & Preckel, 2017, see below for further discussion). This study focuses on the critical stage of childhood (Grade 4), where the formation and differentiation of ASC start influencing further academic performance and enjoyment (Pinxten et al., 2014). Therefore, research on how comparison processes forming ASC function for primary school students and generalize across countries would be practically relevant for teachers' feedback practices and for policy and intervention strategies aiming to promote ASC.

Dimensional and social comparison processes have been well-documented in the I/E model and its extension to DCT, and BFLPE theory. First, we provide an overview of these theoretical models and limitations of previous ASC research on primary school students.

The I/E model and DCT: Dimensional Comparison Effects in Primary-Age Children

The I/E model posits that students form their verbal and mathematic self-concepts as a function of two underlying processes: external comparison, and internal (dimensional) comparison. Students engage in external comparisons by comparing their performances with those of other students, so that comparatively higher mathematic achievement than their classmates should result in higher mathematics self-concept (see matching paths in basic I/E model structure depicted in Figure 1). Students also conduct dimensional comparisons (as mentioned above) - which are ipsative- so that high levels of mathematics ability should result in lower verbal self-concept (negative cross-paths). The I/E model implies that good mathematics achievement would boost one' mathematics self-concept, which companies with a parallel decrease in verbal self-concept caused by dimensional comparison effects. Thus, positive within-domain matching paths in conjunction with negative between-domain cross-paths lead to an increasing differentiation of ASC.

The cross-cultural generalizability of the I/E model was first tested by Marsh and Hau (2004). They used PISA 2000 data and found the external comparison processes (Mean effect sizes [ES] = .51/.47 from mathematic/verbal achievement to mathematic/verbal self-concept), and the dimensional comparison processes (-.22/-.21 from verbal/mathematic achievement to mathematic/verbal self-concept) held invariant across 26 countries (103,558 15-year-old students), providing strong cross-cultural evidence. In a subsequent meta-analysis based on 69 data sets with 125,308 students, Möller et al. (2009) found the I/E associations consistent across ages, measures, gender, and countries. Nevertheless, the meta-analysis sample was dominated by adolescent cohorts, and only 3 out of 69 samples were from children in grade 4 or younger. More specifically, these three primary-school studies did not use representative samples and only covered three countries (Australia, China, and Germany, see Möller et al., 2009). Recently, several individual studies addressed this shortcoming and tested the I/E model based on primary-age cohorts (e.g., Ehm et al., 2014; Lohbeck & Möller, 2017; Schmidt et al., 2017; Pinxten et al., 2015). These studies provided evidence of the I/E predictions for children in Grade 3 and 4 (Ehm et al., 2014; Pinxten et al., 2015) but not in Grade 1 and 2 (Ehm et al., 2014; Lohbeck & Möller, 2017). Again, none of these studies were drawn from nationally representatively sample. Given that dimensional comparison processes play a different role in contributing to self-concept development and differentiation at different cognitive development stages, it is critical to test how the I/E prediction generalize across countries in primary school.

Based on the I/E framework, some studies have applied the I/E model to other subjects, such as native language vs. foreign language (e.g., Xu et al., 2013) and mathematics vs. science (Chiu, 2012; Marsh et al., 2015b). These studies found a pattern similar to the original I/E model. For instance, Marsh et al. (2015b) found consistently positive matching paths (e.g., mathematic achievement predicts mathematic self-concept) and negative non-matching paths (cross-paths, e.g., mathematic achievement to science self-concept) from achievement across countries, but the pattern of results was stronger for eighth-graders (Mean ES = -.269) than for fourth-graders (-.203). This indicates that fourth-graders are able to distinguish between mathematics and science and engage in contrasting dimensional comparisons. Recently, Möller and Marsh (2013) grouped these studies and articulated the development of self-concept in a more general DCT. The DCT extends the I/E model and assumes that dimensional comparison processes in school context tend to involve many domains rather than just mathematics and verbal domains, and how students choose comparison standards is driven by not only motivational needs but also by contexts or norms without intent or awareness (Möller & Marsh, 2013).

Why is it important to incorporate multiple core subject domains to evaluate dimensional comparisons? Dimensional comparisons are a double-edged sword, since they lower the self-concept in the worse off domain while raising it in the better off domain (Möller & Marsh, 2013; Müller-Kalthoff et al., 2017). Thus, the consequence of dimensional comparisons is responsive to the standard selection process – which dimension is chosen as the standard for a particular self-evaluation in a target domain. In the typical I/E literature, researchers constrain their focus on mathematics and verbal domains that are separated by the greatest distance on the continuum of ASC and propose that students are likely to use these two domains as target and standard domains when they evaluate their abilities in either domain. However, evaluating dimensional comparisons based on two domains would result in exaggerated negative contrast effects because the domains selected in the I/E model are implicitly treated as standard and target and thus the differentiation between the domains drives dimensional comparisons, neglecting comparisons with other domains. For example, Xu et al. (2013) found that the negative contrast effects between native language and foreign language disappeared when mathematics was included in the I/E model; instead, the negative effects occurred between mathematics and both verbal domains. As such, inclusion of multiple domains, particularly for core subject domains, will help us better understand how people calibrate their self-concept of abilities in different domains through dimensional comparison processes.

DCT further postulates that ASCs are formed by two dimensional comparisons: (1) contrasting dimensional comparison processes, in which good performance in one domain leads to lower ASC in other domains (i.e., contrast effects); and (2) assimilating dimensional comparison processes, in which good performance in one domain leading to higher self-concept in other domains (i.e., assimilation effects). Whether students engage in contrasting or assimilating dimensional comparisons is related to their beliefs as to how similar or dissimilar different subject domains are to each other (Möller et al., 2015). One of the critical assumptions of DCT is that perceived subject similarity corresponds to the verbal-mathematic continuum of core ASC domains (Möller & Marsh, 2013, see Appendix 1 in Supplement Materials). Haag and Gotz's (2012) supported this assumption and found that subjects far from each other on the continuum (e.g., math vs. German) with low ASC correlations were perceived as rather dissimilar and that subjects (close to each other, e.g., math vs. physics) with high ASC correlations are perceived as more similar. Thus, according to the verbal-mathematical continuum, assimilation effects are assumed to occur between "near" domains (e.g., mathematics vs. physics, native language vs. foreign language), whereas contrast effects are assumed to occur between "far" domains (e.g., mathematics vs. reading). This prediction has been supported in several empirical studies (e.g., Guo et al., 2017; Jansen, Schroeders, Lüdtke, & Marsh, 2015; Möller et al., 2015; Marsh, Kuyper, Morin, Parker, & Seaton, 2014; Marsh, Lüdtke et al., 2015) and a recent experimental study (Helm, Mueller-Kalthoff, Nagy, & Möller, 2016) based on secondary-age cohorts.

A critical limitation of previous cross-cultural studies on the DCT predictions is that only two subject domains (mathematics vs. science, or mathematics vs. reading) are considered. As a consequence, it limits our understanding of how dimensional comparisons among reading, mathematics, and science, which represent a broad spectrum of the verbal-mathematic continuum, predict ASC development and differentiation from a cross-cultural perspective.

The BFLPE: Social Comparison Effects in Primary-Age Children

According to BFLPE, students compare their own academic achievement with the achievements of their classmates and use this social comparison as a basis or reference against which they form their own ASC (Marsh, 2007). BFLPE indicates that students attending mixed- or low- ability schools judge themselves more positively and have higher ASC than comparable students (i.e., equal in ability) attending high achieving schools, which is a negative effect of school-average achievement on ASC.

There is now considerable empirical support for the cross-cultural generalizability of the BFLPE. More specifically, based on 15-year-age students from three successive PISA data collections, Marsh and his colleagues found the negative effect of school-average achievement on ASC in the general academic domain (Marsh & Hau, 2003: 103,558 students from 26 countries, ES = -.20), mathematic domain (Seaton, et al., 2009: 265,180 students from 41 countries, ES = -.49), and science domain (Nagengast & Marsh, 2012: 397,500 students from 57 countries, ES = -.19). Recently, Marsh, Abduljabbar, et al. (2015b) drew on the TIMSS 2007 data and showed significant BFLPE on ASC in mathematics for fourth- and eighth-graders across 13 countries (117,321 students). More importantly, the BFLPE was systematically larger in the eighth-grade cohort (mean ES = -.46) than in the fourth-grade cohort (-.28). Indeed, apart from cognitive development, school environment becomes more achievement-oriented in secondary school, which also leads to increased social comparisons and the BFLPE (Eccles et al., 1993; Hattie, 2012). More recently, the BFLPE has been found to be related with the structure of educational system; the BFLPE is stronger in countries with greater ability stratification (Salchegger, 2016; Parker et al., 2017). In tracked/stratified countries, bright students are likely to be placed in more competitive and higher achieving school; the reverse is true for low-achieving students. Such school composition provides more “distorted” frame-of-reference for social comparisons. It strengthens the BFLPE and leads to more unrealistic ASC for individual students, which might not be reflected by their actual academic abilities (see Parker et al., 2017 for more discussion). Thus, the BFLPE articulates how local school context influences ASC development through social comparison processes.

While the majority of studies on the BFLPE focuses on math, some recent individual studies have also provided the evidence of the BFLPE in verbal domains (native and foreign languages) in secondary school (e.g., Marsh, Kuyper et al., 2014; Parker, Marsh, Lüdtke, & Trautwein, 2013) even in younger age groups such as primary school children (e.g., Lohbeck & Möller, 2017; Pinxten et al., 2015; Roy, Guay, & Valois, 2015). Again, the studies based on young cohorts showed that the BFLPE was relatively small. For example, Roy et al. (2015) found that the effect size of the BFLPE on verbal self-concept was only -.14 based on a sample of Canadian students between 8 and 12 years old. However, investigations of the cross-cultural generalizability of BFLPE in verbal domain for either primary or secondary school students have been scarce.

Recently, several studies have integrated the I/E model and the BFLPE and examined how the dimensional comparison and social comparison processes mutually shape one's ASC using more sophisticated statistical models (Chiu, 2012; Parker et al., 2013; Lohbeck & Möller, 2017; Pinxten et al., 2015; Schurtz Pfost, Nagengast, & Artelt, 2014). These studies consistently found that the effect sizes of the I/E model and the BFLPE are slightly smaller in the integrated ASC model than those in separated models. Some researchers also tested cross-domain paths from school/class-average achievement to students' individual ASC (e.g., class-average mathematic achievement to verbal self-concept) in the integrated model. However, Pinxten et al., 2015 (also see Lohbeck & Möller, 2017) showed that these additional cross-domain paths were trivial in size and ignorable for the primary-age cohort. Again, there is, to our knowledge, neither individual nor cross-cultural studies that have incorporated multiple (more than two) subject domains (i.e., reading, mathematics, science) and examined the integrated ASC model.

The Present Investigation

Drawing on DCT (an extension of the I/E model) and the BFLPE, we aim to examine how social comparison and dimensional comparison processes mutually shape fourth-graders' multidimensional ASCs across 15 OECD countries. Importantly, this study combines two psychological processes in an integrated ASC model and explores the cross-cultural generalizability of the predictions across reading, mathematics, and science. Hence, the present study is unique in that it takes three core academic domains into account and integrates the DCT and BFLPE to provide a greater understanding of the generalized motivational processes forming primary school students' ASCs.

Based on our review, our hypotheses are as follows:

1. **The DCT predictions (H1):** We predict matching paths from each of the three achievement domains (reading, mathematics, and science) to corresponding ASCs will be significantly positive (H1a). According to the verbal-math continuum of ASC, we hypothesize that the cross-paths between mathematics and science will be less negative (or even positive), compared to "far domain" cross-paths between mathematics and reading; in the same vein, cross-paths between reading and science are expected to be less negative (or even positive) compared to those between reading and mathematics (H1b). More specifically, we expect that cross-paths between mathematics and science will be less negative (or more positive) than those between science and reading (H1c), given that mathematics and science are closer to each other in the verbal-math continuum of ASCs.
2. **The BFLPE predictions (H2):** We expect significantly negative effects of class-average achievement on ASCs in the matching domain.
3. **Generalizability of the results (H3):** Cross-cultural comparisons provide researchers with a heuristic basis to test the external validity and generalizability of their measures, theories, and models. Typically, there are two main approaches to cross-cultural comparisons: the universalist and relativist perspectives. The universalist perspective refers to the cultural universals with an emphasis on cross-cultural similarities of theoretical predictions and replicability of results, whereas the relativist perspective refers to phenomena specific to a particular culture with an emphasis the uniqueness of an individual case in its own terms. The present study is unique in that it focuses on both universalist and relativist perspectives. Specifically, the DCT predictions (particularly the cross-paths) is more relevant to self-evaluation based on intraindividual differences in abilities among different subject domains. As such, in line with previous cross-cultural studies (e.g., Guo et al., 2017; Marsh & Hau, 2004),

we predict that the dimensional comparison processes will be invariant across countries (universalist perspectives) (H3a). In contrast, the BFLPE is more related to the influence of macrocontext (i.e., the structure of educational system and school composition), and thus we expect that the BFLPE will generalize across countries and its strength will be associated with the level of ability stratification at country level (Salchegger, 2016; Parker et al., 2017, relativist perspectives) (H3b). Finally, the pattern of these predictions is expected to be weaker than those reported in previous studies focusing on secondary-age cohort, because of younger cohort's cognitive development and environment changes (H3c).

Method

Participants

PIRLS is an international assessment of reading comprehension for nationally representative samples of fourth-grade students from participating countries. TIMSS is an international assessment of mathematics and science competence of fourth- and eighth-grade students. Both are nationally representative samples. In 2011, 34 countries administered the TIMSS and PIRLS assessments to the same samples of fourth-grade students (Martin & Mullis, 2013. In this study, we focused on all OECD countries who participated in both assessments. In total, we considered data from 15 OECD countries with 67,386 students in 3808 classes and 2564 schools (see Table 1) – a ratio of approximately 1.5 classes per school.

Measures

Academic self-concept. Students responded to items designed to measure self-concept in reading (4 items), mathematics (5 items), and science (5 items) on the same classic four-Likert (agree–disagree) response scale; three of the ASC items had the same wording across the three subject domains: (e.g., "I usually do well in reading/mathematics/science", "Reading/Mathematics/Science is harder for me than any other subject"), but there were minor wording changes for the other ASC items. The ASC latent constructs for reading, mathematics, and science demonstrated satisfactory reliability across countries (Cronbach's alpha α/*SD*= .76/.05, .83/.03, and .80/.03 respectively, also see Appendix 2 for specific items and factor structure).

Academic achievement. Two question formats were used in the TIMSS and PIRLS assessments – multiple-choice and written-response questions, to assess participants' academic ability of reading, mathematics and science (Martin & Mullis, 2013). Specifically, science ability is assessed though a range of questions in the three science subdomains (a 45% focus on life science, 35% on Physical science, and 20% on Earth science).

Data analysis

In the present study, multi-group multilevel confirmatory factor analyses (CFAs) and structural equation models (SEMs) were conducted with Mplus 8 using the robust maximum likelihood estimator. To account for a nested data structure in which students are nested within schools and classes, we used the Mplus complex design to control for clustering of students within classes and schools with the HOUWGT weighting variables (see Appendix 8 for an annotated syntax, also see Marsh, Abduljabbar et al., 2015b for more details). In the TIMSS and PIRLS 2011 database, five plausible values were generated for each pupil to estimate their proficiency in each subject. We used full information maximum likelihood (FIML) estimation to handle a relatively small amount of missing data (an average of less than 2.5% per item). To fully account for the plausible values of test scores, analysis had been done separately for each of the five data sets based on different plausive values and then combined the results using the Rubin (1987) strategy to obtain unbiased parameters estimates, standard errors, and goodness-of-fit (See Appendix 3).

We used a latent-manifest (i.e., latent variable with manifest aggregation) approach (Marsh et al., 2009) in which multiple indicators are used to infer Level 1(L1: student level) and Level 2 (L2: class level) latent ASC factors, whereas manifest aggregation was used to form the L2 achievement. For TIMSS and PILRS data, intact classes were sampled so that the sampling ratio approached one, and thus sampling error was minimal. With this strategy, the use of latent aggregation for L2 achievement can overcorrect BFLPE estimates (see Marsh et al., 2009; Marsh, Abduljabbar et al., 2015a for further discussion). Test scores were grand-mean centered rather than group-mean centered at L1. As such, in this model estimates of L2 effects are already controlled for L1 effects, meaning that the BFLPE is defined as the L2 path coefficient (β, from achievement to ASC in the matching domain). ESs for the BFLPE were calculated according to the recommendations by Marsh, et al. (2009), with the following formula: $ES = 2 \times \beta \times \frac{SD_{predictor}}{SD_{outcome}}$ using the "Model Constraint command" in Mplus (See Appendix 8). We used the delta method to calculate summary statistics (e.g., Mean[M]) of effect sizes across countries.

Preliminary analyses

Preliminary analyses, detailed in Appendix 4, demonstrated: (a) there was good support for the factor structures underlying mathematics, science, and reading self-concepts, based on multilevel-multigroup analyses (CFI = .955, TLI = .946, RMSEA= .036); (b) factor loading invariance for the three ASC constructs was achieved across L1 and L2 levels as well as the 15 countries (CFI = .950, TLI = .945, RMSEA= .036). This constrained model was the basis of subsequent results.

Results

Correlation between Student Level (L1) Achievement and Self-concept

On average, reading, mathematic, and science achievements are highly correlated (*r* = .770 to .854), whereas the correlations among the three ASCs were substantially smaller across countries (*r* = .249 to .374, see Table 2). For reading and mathematics domains, correlations between ASCs and the matching achievement were moderate (*r* =.399 and .314, respectively) and slightly larger than those between each ASC and the non-matching achievement (r = .142 to .338). However, within-domain correlation between ASC and achievement was somewhat weaker for science (r = .202), which was even smaller than some cross-domain correlations (e.g., between reading self-concept and science achievement, .338). In summary, results indicated good support for the high domain specificity of the three ASCs and reasonable support for convergent and discriminant validities of ASCs in relation to achievement for mathematics and reading, but not for science.

Tests of Prediction Relating to the DCT In the Integrated Model

In the SEM model, we included 9 (3 x 3; 1 matching path + 2 non-matching paths for reading, mathematics, and science) paths from L1 achievement in each domain to each of the three L1 ASCs to test the predictions of the I/E model, as well as 3 matching paths from L2 achievement to L2 (latent aggregation) ASCs to test the predictions of the BFLPE. Note that relations from the three L2 achievements to non-matching L2 ASCs (e.g., L2 mathematics achievement and L2 reading/science self-concept) were represented as correlations in this model[2]. This integrated model showed a good model fit (CFI = .944, TLI = .926, RMSEA= .047, see Appendix 6 for results based on models including all possible pairs of the two subject domains [e.g., mathematics vs. reading, reading vs. science]).

The pattern of the I/E predictions (9 paths) are presented in Figure 2 (also see Appendix 5 for country-specific and averaged path coefficients). The matching paths from achievement to ASCs were substantially positive for reading (Mean [M] = .422, SE = .010)

[2] Previous studies suggest that it is also important to examine L2 cross-domain BFLPE patterns in the integrated ASC model (e.g., Parker et al., 2013). In supplemental analyses, we replaced correlations from the three L2 achievement to non-matching L2 ASCs as paths in the SEM (in total 6 paths), which resulted in the same model fit as the original SEM model. Results indicated that the coefficients of the 6 non-matching L2 paths were trivial in size across countries. Subsequently, we tested a restricted model where the 6 non-matching L2 paths were constrained to be zero. For the TLI and RMSEA that control parsimony, the fit of the restricted model became even better (ΔTL1 = .002; ΔRMSEA = -.001) and indicated that the cross-domain BFLPEs were ignorable, which is consistent with previous findings based on primary students (e.g., Pinxten et al., 2015; Lohbeck & Möller, 2017).

and mathematics (*M* = .422, SE = .008), but were somewhat weaker for science (*M* = .191, SE = .011) across countries. In relation to non-matching paths, slightly negative or non-significant paths from mathematic achievement to reading self-concept (*M* = -.064, SE = .009) and from reading achievement to mathematic self-concept (*M* = -.088, SE = .010) were present in each country. Small negative non-matching paths were found from mathematic achievement to science self-concept (*M* = -.087, SE = .010), whereas the paths from science achievement to mathematic self-concept were insignificant or slightly positive across countries (*M* = .032, SE = .010). The non-matching paths from read achievement to science self-concept are consistently positive (*M* = .140, SE = .009), while the paths from science achievement to reading self-concept were insignificant or slightly positive across countries (*M* = .035, SE = .011, see subsequent discussion).

In summary, as expected, the matching paths were significantly positive and stronger than the cross-paths (H1a), particularly for mathematics and reading. Also, aligned with Hypothesis H1b, on average the cross-paths between mathematics and reading were more negative than those between mathematics and science as well as between science and reading (see Figure 2 and Table S5A). However, we found positive cross-paths between science and reading, but negative cross-paths between mathematics and science, which are inconsistent with our expectations (H1c).

Tests of prediction relating to the BFLPE in the integrated Model

Across all countries, class-average achievement had a strong and negative effect on ASCs in the matching domain for mathematics (*M* = -.323, SE = .010), followed by reading (*M* = -.222, SE = .010, see Figure 3, see Appendix 5 for more details). However, the BFLPE was somewhat smaller for science across countries (*M* = -.161, SE = .017), and was not significant for 6 of the 15 countries. Overall, there is good support in relation to the BFLPEs for the aggregate across countries for all three domains (H2). The pattern of results is consistent for each of the countries considered separately for mathematics and reading, but not for science.

Generalizability of the DCT and the BFLPE over countries

The direction and effect sizes of the predictions in relation to the DCT and the BFLPE are largely similar over 15 OECD countries, which supports the a priori prediction (H3). As seen in Figure 2, the pattern of the DCT predictions in relation to mathematics and reading self-concepts was predominately aligned across countries. Specifically, for the effects on mathematic self-concept, mathematic achievement is the strongest and positive predictor, followed by science achievement (near-zero effect), and then reading achievement (slightly negative effect). For the effects on reading self-concept, reading achievement is the strongest and positive predictor, followed by science achievement (near-zero effect), and then mathematic achievement (slightly negative effect). However, science achievement was not consistently found to be a stronger positive predictor of science self-concept than reading achievement (only five countries, see Figure 3), and both achievements tended to have similar effect sizes. Mathematic achievement was consistently shown to negatively predict science self-concept. For the pattern of the BFLPE, all countries consistently showed stronger BFLPE in mathematics than in reading and science.

To more directly compare the similarity of country-specific path coefficients, we also calculated a profile similarity index (PSI). The PSI is an estimate of the correlations between path coefficients obtained from different countries. In the multiple-group integrated ASC model there were 12 path coefficients (9 paths relating to the DCT and 3 paths relating to the BFLPE) for each country. We calculated the correlations between these path coefficients across the 15 countries. The PSI indicated the high level of similarity across countries (range from 0.76 to 0.99). Thus, there is good support for the generalizability of path coefficients over 15 countries. More specifically, in line with our expectation (H3a), between-country variance for dimensional comparisons (cross-paths) was relative small (*SD* < .08), and the imposition of the constraints on all 6 cross-paths across countries resulted in negligible changes in model fit (ΔCFI = .001, ΔTLI = -.001, equal RMSEA). We employed intracluster correlations (ICCs) for achievement by schools to measure the degree of ability stratification at country level (Parker et al., 2017) and to test the relationship between the structure of educational system and the BFLPE. We found that mathematics ICC was negatively and highly correlated with the BFLPE (r = -.679, see Appendix 7), whereas the pattern of results for reading and science was somewhat weaker (r = -.312 and -.241, respectively). Consistent with a priori predictions (Hypothesis H3b), these findings demonstrate that the BFLPE is stronger in educational systems that are more highly stratified by achievement.

Comparisons to previous studies on adolescence. The effect sizes of the cross-paths (dimensional comparisons) between mathematics and reading in the current study were smaller compared to those reported in previous cross-cultural studies and meta-analysis (Marsh & Hau, 2004; Marsh, Abduljabbar, et al.2015a, Möller et al., 2009) primary based on secondary-age cohorts. For example, the mean of the paths from mathematics achievement to reading self-concept was -.081, which is weaker than that reported in Marsh and Hau's (2004) study in OECD countries (-.21) and Möller et al's meta-analysis (-.21) (similar pattern was found in the model where only mathematics and reading were included, Appendix 6). For the BFLPE, the effect sizes in the current study were smaller for science and particularly for mathematics, compared to those reported in previous cross-cultural studies based on secondary-age cohorts (Seaton et al., 2009; Nagengast & Marsh, 2012; Marsh, Abduljabbar, et al.2015b). For example, the mean of the BFLPE for mathematics was -.323, which is weaker than that reported in Marsh, Abduljabbar, et al.'s (2015b) study based on eighth-graders (-.463) in TIMSS 2007. Given that this is the first large-scale cross-cultural study examining the BFLPE in a verbal domain, we are only able to compare the effect size with other individual studies targeting secondary-age cohorts. Again, the BFLPE for reading (-.222) is smaller than that in the studies drawn from lower-secondary (-.271, Marsh, Kuyper et al, 2014) and upper-secondary cohorts in general schools (-.41, Parker et al., 2013). Thus, in line with our hypotheses (H3c), the pattern of results in relation to dimensional and social comparisons are consistently smaller than that based on secondary-age cohorts.

Discussion

The results of the present investigation are largely consistent with our anticipations: primary school students are likely to engage in the generalized motivation process (both dimensional and social comparisons) to form their reading, mathematic, and science self-concepts, providing good support for cross-national generalizability. This is important because this is the *only* large-scale cross-cultural study to integrate the I/E model and the BFLPE across the three core subject domains.

Our findings largely support the crucial assumption of DCT that students tend to make both assimilating and contrasting dimensional comparisons based on the verbal-mathematics continuum of ASCs. Students are likely to engage in contrasting dimensional comparison between mathematics and reading, being dimensions on opposite ends of the ASC continuum; and this contrasting dimensional comparison is stronger than that between mathematics and science, and between science and reading. While science is more closely related to the mathematical side than to the verbal side of the ASC continuum, students are likely to engage in assimilating dimensional comparisons between science and reading but contrasting dimensional comparisons between mathematics and science. This suggests that students apparently perceive science and reading to be similar and complementary. In particular, significant and positive effects from reading achievement to science self-concept suggests that reading achievement is an important subject for students to evaluate their strengths and weaknesses in science. A possible justification for this phenomenon is the typical science curriculum structure in primary schools. The science curriculum aims to develop elementary knowledge of life (e.g., plant and animal structure, life processes and cycles), physical (e.g., some properties of matter, electricity, and energy), and earth (e.g., solar system, Earth's physical characteristics and resources) sciences, as well as cultivate skills to demonstrate this elementary knowledge by providing brief descriptive (written) responses combining of science concepts with information from both tangible, and abstract contexts. As such, science learning in primary schools involves heavy reading and writing components. In contrast, the typical mathematics curriculum is more focused on developing skills in solving problems involving operations with numbers (e.g., multiplication and division), line symmetry and geometric properties, as well as tables and graphs (e.g., pictographs and tally charts). Furthermore, the characteristics of the test items in TIMSS2011, where the nature of achievement tests is more closely related to the academic curriculum than in PISA (see Marsh, Abduljabbar et al., 2015b for further discussion), also reflect much higher reading demand for science (168 items) than for mathematics (175 items). On average, the total number of words (per item) was substantially lower for mathematics than for science (25 vs. 41), whereas mathematics items have more symbolic language (e.g., numerals and operators) and data displays (e.g., geometric shapes and graphs) when compared to the science curriculum (5 vs. 1 and 8 vs. 3 respectively; Martin & Mullis, 2013).

The contrasting dimensional comparison between mathematics and science might be a result of the choice of comparison standards driven by the context or norms (Möller & Marsh, 2013). A recent experimental study (Helm & Möller, 2016) found that the pattern of the DCT predictions was related to the basis of assessment of students' ASCs. Specifically, based on a sample of German 5th-12th graders, Helm and Möller (2016) showed that measuring German and mathematic self-concepts at the same time – the norm for I/E research – strengthened contrast effects between these two domains. Such contrast effects became much weaker when German and mathematic self-concepts were assessed in separate questionnaires. In this study, the fourth-graders participating in both TIMSS and PIRLS completed the mathematics and science module together, but the reading module was conducted separately. When mathematics and science self-concepts are assessed together, participants are more likely to implicitly choose mathematics/science as the standard domain for dimensional comparisons, which may amplify the contrasting comparison process (see Möller & Marsh, 2013 for further discussion). This also helps explain the findings in previous cross-cultural studies with evidenced contrast I/E predictions between mathematics and science without controlling for reading achievement, based on the TIMSS 2003 and 2007 assessments (Chiu, 2012; Marsh, Abduljabbar et al., 2015a). We note that our study is the first large-scale analysis to incorporate the reading domain based on primary-age children. Thus, the results provide new theoretical and substantive insights into DCT predictions.

Of particular relevance, nearly all BFLPE studies have been based on a single domain, and only several recent studies have considered two domains using data from individual countries (e.g., Marsh, Kuyper et al., 2014; Parker et al., 2013; Pinxten et al., 2015). Apparently, no research—and particularly no cross-cultural research—has been done with three core academic domains based on nationally representative samples of students. Thus, the sizes of the BFLPE across domains has rarely been directly compared. In this respect, the combined TIMSS and PIRLS data are ideally suited to evaluating the juxtaposition of nationality and multiple subject domains. Consistent with previous individual studies (e.g., Marsh, Kuyper et al., 2014; Parker et al., 2013; Pinxten et al., 2015), the sizes of the BFLPE are stronger in mathematics than in reading. A potential reason is that reading skill is necessary in multiple aspects of daily life and is more closely related to the social context outside of school. The formation of reading self-concept is thus more likely to be based on students' experience within, but also outside the school environment. As such, students' social frames-of-reference for reading may be more complex, more general, and more indicative of a broader context than for mathematics. Indeed, researchers have found strong BFLPEs in foreign language (i.e., English as a second language), because students are less likely to have contexts outside of school for assessing their performance (see Parker al., 2013; Schurtz et al., 2014 for further discussion). When students' opportunities to evaluate their relative competence is restricted to their school environments, school average achievement will provide a pure measure of the social comparative environment, and therefore we would expect BFLPEs to be larger.

Consistent with a previous cross-cultural BFLPE study on secondary school students (Nagengast & Marsh, 2012), the BFLPE on science is consistently smaller than that on mathematics. As mentioned above, one of the possible reasons is that science learning in primary schools heavily involves reading components and thus, social comparison processes for science may be more related to a broader context outside of school. Moreover, the correlation between self-concept and achievement in science is only modest (M = .202), which is one of the bases underlying social comparison mechanisms. This modest correlation may be due to the weak linkage between the content taught in primary classes and that tested in the TIMSS assessment. Compared to mathematics and native language curricula, there is not a uniform science curriculum for children; and teachers who usually instruct multiple subjects in primary schools do not have extensive scientific training before they begin teaching science (see Mullis et al., 2012 for the summary of different science curricula across countries).

It should be noted that external comparisons (external frame-of-reference) in the I/E model are often referred to as social comparisons in the I/E studies, which causes that some ASC researcher conflate it with social comparisons posited in the BFLPE. However, both social comparison processes theoretically operate in different ways. External comparison refers to the information available in one's environment that individuals use as a basis for self-evaluation, including the perceived abilities of others, and external achievement feedback (e.g., grades and test scores). In relation to the I/E model and DCT, the external comparisons are in relation to external information about student achievement (e.g., grades and test scores) at the level of the individual student. As such, the higher achievement feedback a student gets, the higher self-concept he/she has, which leads to positive external (social) comparison processes. On the other hand, social comparison in the BFLPE is a contextual effect on the school/class level, net of the effects of external comparison at the individual level. Thus, in relation to the BFLPE, the social comparisons are in relation to school or class-average achievement. The BFLPE presents a counterfactual case that a student placed in a high-ability school will have lower self-concept than if they had been placed in a lower-ability school. This is because social comparison is highly related to their relative ability rank order positioning within a student's specific school. In a high-ability school, a student is more likely to be ranked among the low achieving students than in a low-ability school, which results in negative social comparison processes. Therefore, social comparisons posited in the I/E models and BFLPE refer to different levels of self-evaluation mechanisms mutually shaping ASCs.

Cross-cultural Generalizability and Its Implications

Our findings provide support for cross-cultural generalizability of the DCT and BFLPE predictions, which reflects a tendency of students to use their performances in different domains (dimensional comparison), and local comparison groups (social comparison), as reference points to base self-evaluation. Yet, the effect sizes for these predictions were consistently lower and some directions were even reversed, compared to those obtained from similar studies with adolescents, due to younger children's comparatively lower cognitive level, limited learning experience, and less competitive learning environment. Importantly, this study explores dimensional comparisons in relation to science domain, which has received scant attention in previous ASC research focusing on primary school students. Results show that students are likely to engage in positive assimilating comparisons between science and reading rather than negative contrasting comparisons (evident in ASC research based on secondary school students, see Jansen et al., 2015). These findings indicate that dimensional and social comparisons operate in different ways in contributing to ASC development and differentiation at different development stages. Hence, more effort should be devoted in the future to study ASC development of young children by incorporating multiple subject domains and multiple ASC theoretical models, given that empirical ASC studies based on adolescent cohorts have dominated the literature. Specific implications for instructional practices were as follows.

The dimensional comparison processes contribute to the differentiation of ASC across domains in primary schools. For example, the contrast dimensional comparisons between mathematics and reading was associated with low correlation between ASCs in the two domains (.256). Such ASC differentiation will lead to further distinction between performance in both domains, which in turn drives the dimensional comparisons (the reciprocal I/E model, Möller, Retelsdorf, Köller, & Marsh, 2011). Importantly, despite substantial variations in reading and science curricula (see footnote 1 and Mullis et al., 2012), the pattern of dimensional comparisons is invariant across countries, supporting the external validity of our findings. However, dimensional comparisons do not seem to be carried out by teachers when they estimate students' ASC (Helm, Müller-Kalthoff, Mukowski, & Möller, 2018); they tend to believe that students who are capable in one academic domain tend to be seen as having high ASC in all domains, while students who are not capable in one area are seen as having low ASC in all domains (Marsh, 2007). This distorted perception is more prevailing in primary school teachers who generally instruct students across multiple school subjects (Marsh, Abduljabbar et al., 2015a for further discussion). As such, they would not be able to provide appropriate feedback and adapt their teaching to meet the motivational requirements of their students. Ironically, awareness of contrast effects between mathematics and reading is imperative for teachers because contrast dimensional comparisons would lead to a large intraindividual difference in achievement and self-beliefs, which in turn constrains students' pursuit in certain educational and occupational pathways at young age (Guo, Parker, Marsh, & Morin, 2015). If teachers better understood formation of ASC in different domains, they would be able to help students adjust the perceptions of subject (dis)similarity, as the contrast effects seem to depend on students' beliefs about the association between mathematical and verbal abilities (Möller & Marsh, 2013). For example, it could be beneficial to show students quite explicitly the similarities between different school subjects—attribution of achievement between every subject regarding interest, effort, and learning strategies (Helm et al., 2016). Thus, the generalized pattern has fundamental implications about the way teachers give feedback to students in different academic domains.

Moreover, our findings greatly expand the scope of support for the generalizability of the BFLPE by combining TIMSS and PIRLS data. The results indicate that the negative BFLPE is prevalent in primary schools where ability tracking or stratification is not implemented in most of education systems. The strong relationship between ICC for schools by achievement and country-by-country variance of the BFLPEs, particularly for mathematics[3], demonstrates that how the makeup of the local school context considerably affects the formation and development of ASC even for primary-age cohorts. On the one hand, in a country with greater ability stratification, students in a low-achieving school are

---

[3] As mentioned above, social comparison processes for reading and science may be more related to a broader context outside of school, which attenuates the correlation between ICC and the BFLPE.

more likely to gain and develop high ASC from the BFLPE. However, these children are difficult to translate high self-beliefs into higher educational performance and ambitious choices, due to the signaling massage that school sends students (e.g., reputation for underperformance and lower selective high school acceptance rate, see Parker et al., 2017). On the other hand, grouping high ability children in the same school will suppress their ASC and thus make less ambitious educational and career decisions than they might otherwise do (Nagengast & Marsh, 2012). Moreover, based on a representative sample of American primary school students, Dicke et al. (2018) further revealed that being placed in a high achieving group of students has a negative impact on a student's ASC, and no positive or even slightly negative effect on their achievement as well. Chiu, Chow, & Joh (2017) also showed that placing children in primary classrooms with high heterogeneity in terms of family background, and past achievement, benefits their overall reading achievement. Therefore our findings, in conjunction with recent BFLPE studies, suggest that social comparison processes are heavily dependent on not only children's cognitive development but also contextual factors and provide a useful perspective for policy makers to construct schools and classes.

Limitations and Directions for Future Research

Several limitations to this study, and some caveats, must be noted. First, as previous cross-cultural studies on the DCT and BFLPE predictions have done, the current study drew on cross-sectional data. Thus, the robustness of our findings could be strengthened by carefully constructed longitudinal panel studies or quasi/true experimental studies to better understand the causal mechanisms between achievement and ASCs. Second, it should be noted that generalizability over 15 OECD countries may not generalize to non-Western contexts. Indeed, Marsh, Abduljabbar et al., (2015a, 2015b) showed that the DCT and BFLPE patterns were smaller in Islamic countries than other countries. We call for research that includes culturally diverse countries (e.g., including Asian and Islamic countries) to evaluate the integrated ASC model using multiple subject domains. Lastly, the dimensional and social comparison processes have been linked to broader motivational consequences (i.e., intrinsic and utility values, Guo et al., 2017; Möller et al., 2015; Schurtz et al., 2014). Therefore, further research is needed to better understand the cross-cultural generalizability of the DCT and BFLPE in relation to the formation of other motivational constructs. Third, our findings in the pattern of dimensional comparisons suggest that how primary school students may perceive similarity of subject domains are quite different from how their secondary school counterparts do. Although previous experimental studies suggesting that lower perceived subject similarity would lead to stronger ASC differences than did higher perceived similarity (e.g., Helm et al., 2016), none of them focuses on primary school students. It, therefore, would be another avenue for future research to examine relationships between perceived subject similarity and dimensional comparisons across subjects on ASC for younger cohorts.

References

Chiu, M. S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology*, 104, 87–107.

Chiu, M. M., Chow, B. W., & Joh, S. W. (2017). Streaming, tracking and reading achievement: A multilevel analysis of students in 40 countries. *Journal of Educational Psychology*, 109(7), 915–934.

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*. Online Advanced Publication.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. Mac. (1993). Development during adolescence. *American Psychologist*, *48*(2), 90–101.

Ehm, J. H., Lindberg, S., & Hasselhorn, M. (2014). Reading, writing, and math self-concept in elementary school children: Influence of dimensional comparison processes. *European Journal of Psychology of Education*, *29*(2), 277–294.

Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2015). Directionality of the Associations of High School Expectancy-Value, Aspirations, and Attainment: A Longitudinal Study. *American Educational Research Journal*, *52*(2), 371–402.

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction*, *49*, 81–91.

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., … Trautwein, U. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open*, *2*(1), 1–20.

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*(8), 1163–1176.

Haag, L., & Götz, T. (2012). Mathe ist schwierig und Deutsch aktuell: Vergleichende Studie zur Charakterisierung von Schulfächern aus Schülersicht [Math is difficult and German up-to-date: Comparative study of characterizations of school subjects from students' points-of-view]. *Psychologie in Erziehung und Unterricht*, 59, 32–46.

Harter, S. (2012). *The construction of the self: A developmental perspective. Vol. 2.* New York: Guilford.

Helm, F., & Möller, J. (2016). Der Einfluss von Erhebungsbedingungen auf Kontrasteffekte dimensionaler Vergleiche. *Psychologie in Erziehung Und Unterricht*, *84*(1), 24.

Helm, F., Müller-Kalthoff, H., Mukowski, R., & Möller, J. (2018). Teacher judgment accuracy regarding students' self-concepts: Affected by social and dimensional comparisons? *Learning and Instruction*, *55*, 1–12.

Helm, F., Müller-Kalthoff, H., Nagy, N., & Möller, J. (2016). Dimensional comparison theory: Perceived subject similarity impacts on students' self-concepts. *AERA Open*, 2(2), 233285841665062.

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. *Journal of Educational Psychology*, *107*(4), 1086–1101.

Lohbeck, A., & Möller, J. (2017). Social and dimensional comparison effects on math and reading self-concepts of elementary school children. *Learning and Individual Differences*, *54*, 73–81.

Marsh, H.W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology.* Leicester, UK: British Psychological Society.

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., … Abu-Hilal, M. M. (2015a). The internal/external frame of reference model of self-concept and achievement relations: Age-Cohort and cross-cultural differences. *American Educational Research Journal*, *52*(1), 168–202.

Marsh, H. W., Abduljabbar, A. S., Morin, A. J. S., Parker, P. D., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015b). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, *107*(1), 258–271.

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376.

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-Cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67.

Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction*, *35*, 16–32.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-Latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764–802.

Marsh, H. W., Pekrun, R., Murayama, K., Katrin Arens, A., Parker, P. D., Guo, J., & Dicke, T. (2017, November 27). An integrated model of academic self-concept development: academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*.

Martin, M. O. & Mullis, I. V. S. (Eds.) (2013) *TIMSS and PIRLS 2011: Relationships Among Reading, Mathematics, and Science Achievement at the Fourth Grade—Implications for Early Learning.* Chestnut Hill, MA: TIMSS & PIRLS.

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H.W. (2015). Dimensional comparisons and their consequences for self-concept, motivation, and emotion. *International Encyclopedia of the Social & Behavioral Sciences*, 26, 430-436

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120,544-560.

Möller, J., Pohlmann, B., Köller, O., Marsh, H. W., Moller, J., & Koller, O. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, *79*(3), 1129–1167.

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48(6), 1315–1346.

Mullis, I.V.S., Martin, M.O., Minnich, C.A., Stanco, G.M., Arora, A., Centurino, V.A.S., & Castle, C.E. (2012). *TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science, Volumes 1 and 2.* Chestnut Hill, MA: TIMSS & PIRLS.

Müller-Kalthoff, H., Jansen, M., Schiefer, I. M., Helm, F., Nagy, N., & Möller, J. (2017, February 6). A double-edged sword? On the benefit, detriment, and net effect of dimensional comparison on self-concept. *Journal of Educational Psychology*. Advance online publication.

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104*(4), 1033–1053.

Parker, P. D., Marsh, H. W., Guo, J., Anders, J., Shure, N., & Dicke, T. (2017). An information distortion model of social class differences in math self-concept, intrinsic value, and utility value. *Journal of Educational Psychology*. Online Publication.

Parker, P. D., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2013). Differential school contextual effects for math and English: Integrating the big-fish-little-pond effect and the internal/external frame of reference. *Learning and Instruction*, *23*, 78–89.

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, *84*(1), 152–174.

Pinxten, M., Wouters, S., Preckel, F., Niepel, C., Fraine, B. De, & Verschueren, K. (2015). The formation of academic self-concept in elementary education: A unifying model for external and internal comparisons. *Contemporary Educational Psychology*, *41*, 124–132.

Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, *42*, 110–116.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons.

Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the big-fish-–little-pond effect across cultures. *Journal of Educational Psychology*, *108*(3), 405–423.

Schmidt, I., Brunner, M., & Preckel, F. (2017). Effects of achievement differences for internal/external frame of reference model investigations: A test of robustness of findings over diverse student samples. *British Journal of Educational Psychology*, 1–16.

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41.

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*(2), 403–419.

Wigfield, A., Tonks, S., & Klauda, S. L. (2016). Expectancy-value theory. In K. R.Wentzel, & D. B. Miele (Eds.), *Handbook of Motivation at School* (pp. 55e74). New York: Routledge.

Xu, M. K., Marsh, H. W., Hau, K.-T., Ho, I. T., Morin, A. J. S., & Abduljabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction. *Journal of Educational Psychology*, *105*(2), 489–503.

Table 1
Demographics

	Sample size			Mean (Test scores)			ICC (Test scores)		
	students	classes	schools	Read	Math	Science	Read	Math	Science
AUS	5943	438	280	509	511	510	.227	.267	.260
AUT	4587	276	158	511	503	525	.130	.195	.166
CAN	4142	220	190	525	526	512	.136	.188	.168
CZE	4433	235	177	535	513	536	.174	.207	.153
DEU	3928	205	197	527	526	525	.201	.192	.197
ESP	4105	200	151	499	484	506	.215	.224	.215
FIN	4541	267	145	555	543	567	.116	.143	.129
HUN	5149	249	149	530	518	538	.260	.300	.307
IRL	4383	220	150	537	524	510	.153	.181	.207
ITA	4125	239	202	525	502	519	.194	.268	.290
NOR	3054	197	119	491	493	492	.098	.165	.137
POL	4962	257	150	513	482	505	.118	.145	.125
PRT	3991	240	147	528	530	518	.147	.261	.208
SVK	5561	314	197	519	503	527	.273	.350	.346
SWE	4482	251	152	523	498	524	.185	.180	.218

Note. Countries represented using International Organization for Standardization three-letter codes.

Table 2

Correlations (MEAN and SD) among Self-Concept and Achievement in Reading, Math, and Science across Countries

	Self-concept			Achievement		
	Reading	Math	Science	Reading	Math	Science
Self-concept						
Reading	-					
Math	.256(.064)	-				
Science	.374(.070)	.249(.061)	-			
Achievement						
Reading	.399(.051)	.214(.060)	.201(.062)	-		
Math	.286(.047)	.314(.060)	.142(.062)	.770(.044)	-	
Science	.338(.053)	.243(.064)	.202(.057)	.854(.032)	.820(.037)	-

Note. Correlation coefficients involving two constructs in matching domains are shaded.

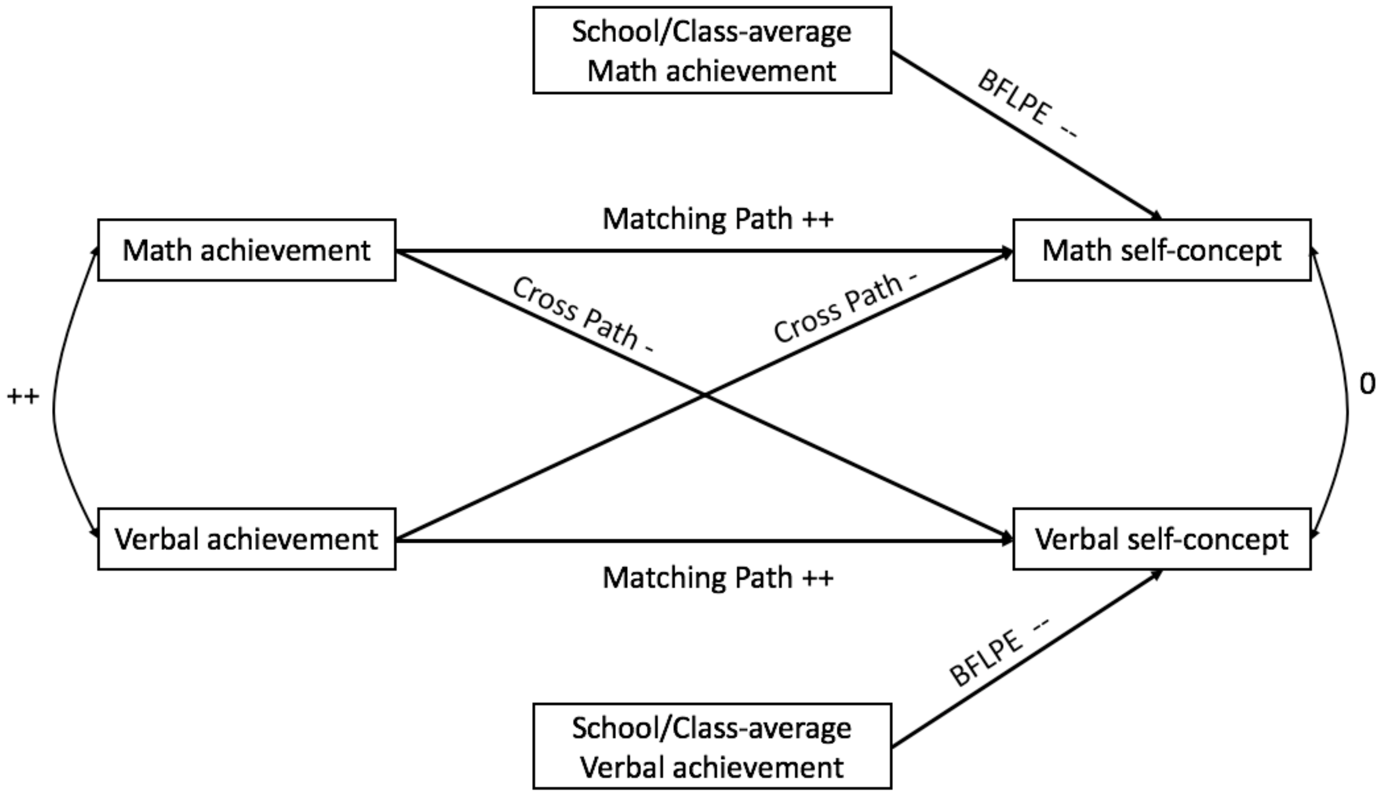

Figure 1. The integrated ASC model formed by the I/E model and DCT.
Some empirical studies on the integrated ASC model also tested paths from school/class-average achievement to students' ASC in non-matching domains (cross-domain BFLPE; e.g., path from class-average math achievement to verbal self-concept). However, the pattern of these new predictions in the integrated models is inconsistent across studies and found to be trivial in size in primary-age cohorts (see Pinxten et al., 2015, Lohbeck & Moller, 2017).

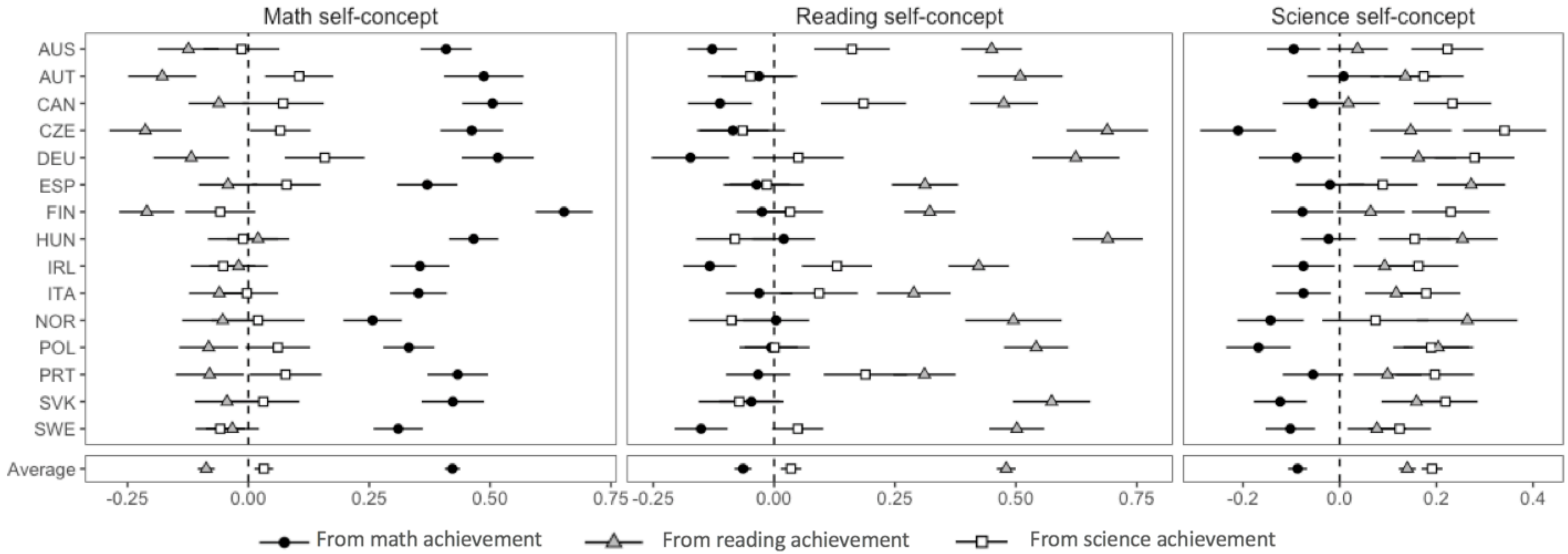

Figure 2. The effects of academic achievement to self-concept in mathematics, reading, and science.
Circles, triangles, and squares indicate paths from academic achievement in mathematics, reading, and science to three domain-specific self-concepts, respectively. The bars indicate the 95% confidence intervals. Countries are represented using the International Organization for Standardization three-letter codes.

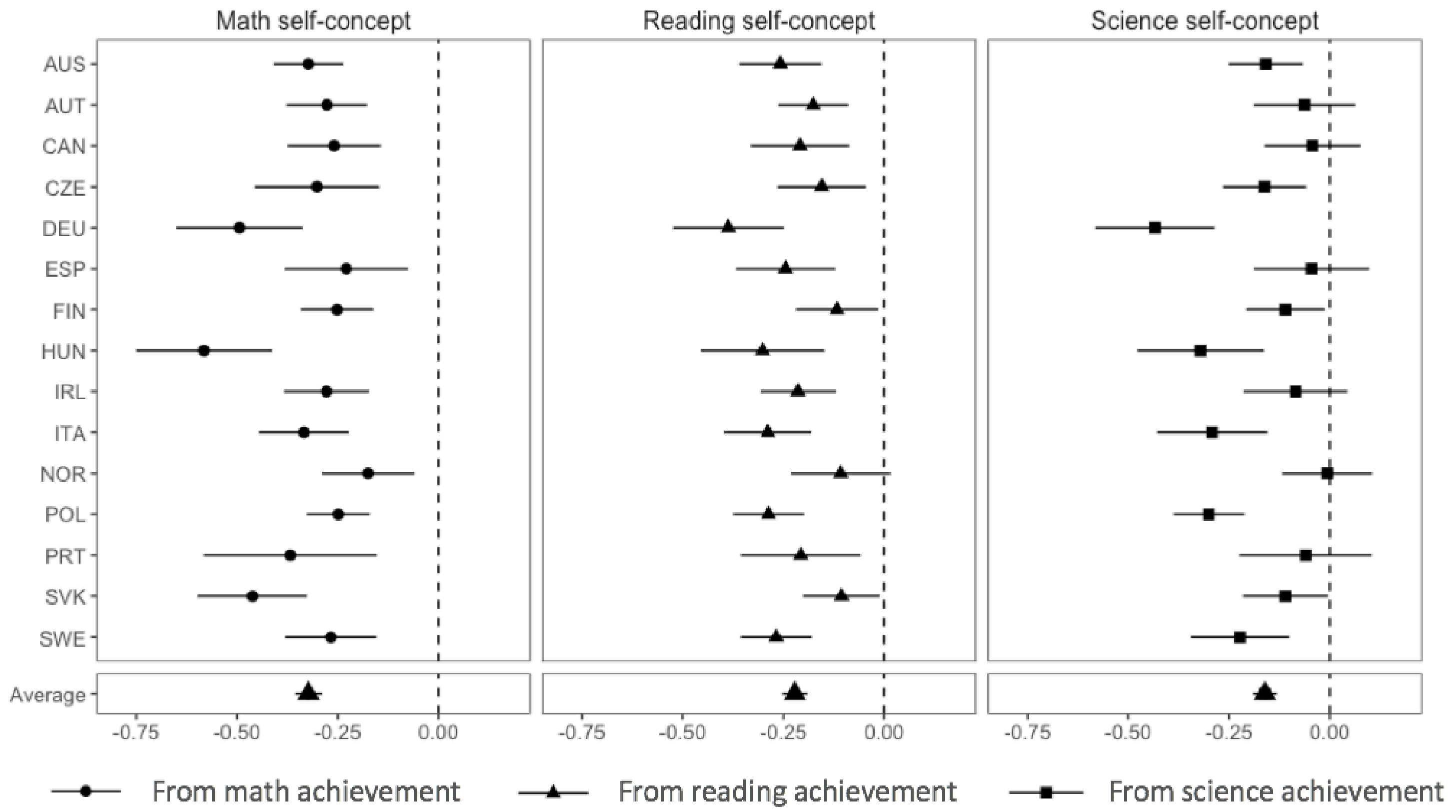

Figure 3. BFLPEs in relation to three subject domains (mathematics, reading, and science).
Circles, triangles, and squares indicate BFLPEs in mathematics, reading, and science, respectively. The bars indicate the 95% confidence intervals. Countries are represented using the International Organization for Standardization three-letter codes.