SCIENTIFIC STUDIES OF READING
2023, VOL. 27, NO. 4, 375–392
https://doi.org/10.1080/10888438.2023.2174028

Routledge
Taylor & Francis Group

The Big-Fish-Little-Pond Effect for Reading Self-Beliefs: A Cross-National Exploration with PISA 2018

Geetanjali Basarkod [a], Herbert W. Marsh [a], Jiesi Guo [a], Theresa Dicke [a], Kate Xu [b], and Philip D. Parker [a]

[a]Institute for Positive Psychology and Education, Australian Catholic University, North Sydney, Australia; [b]Faculty of Educational Sciences, Open University of the Netherlands, Netherlands

ABSTRACT

Purpose: Past research shows the Big-Fish-Little-Pond Effect (BFLPE; negative effect of school-average achievement on student-level academic self-concept) to generalize across countries. However, such evidence is largely limited to math and science. Given that reading self-concept is highly differentiated from math and science self-concepts and plays an important role in predicting educational outcomes, it is essential to examine the universality of the BFLPE and its underlying social-comparison process within this domain.

Method: We assess the cross-national generalizability of the BFLPE for 15-year-olds' reading self-concept using Programme for International Students Assessment 2018 (533,165 students, 72 countries). To demonstrate that the BFLPE operates with a relative—rather than absolute—frame of reference for comparison, we juxtapose difficulty experienced with reading in general (self-concept perceived difficulty; relative frame of reference), with difficulty experienced with reading specifically during the PISA test (PISA test difficulty; absolute frame of reference).

Results: Our findings show that the BFLPE for both the reading self-concept perceived competence and difficulty subscales was robust across countries. Further, the BFLPE was strong for self-concept subscales, but very weak for the PISA test difficulty scale.

Conclusions: Our findings extend support for the generalizability of the BFLPE to reading self-concept and highlight the role of social comparison processes underlying this effect.

Academic self-concept is an individual's perception of their academic abilities. It is shaped by their experiences with and interpretations of their academic performance, and by others' evaluations of their abilities (Marsh & Seaton, 2015). It positively impacts academic achievement (Marsh & O'mara, 2008), interest and satisfaction in school (Marsh et al., 2005), course selection (Marsh, Van Zanden, et al., 2019; P. D. Parker et al., 2014), persistence, and long-term attainment (Guo, Marsh, et al., 2015; Guo, Parker, et al., 2015), above and beyond the effects of prior achievement and IQ. Owing to its importance, it is crucial to understand how people form their self-concepts. Numerous models have been proposed. The common underlying theme in these models is that self-concept is impacted by frames of reference; self-evaluations of competence in a particular domain can be made against many standards of comparison (Marsh & Seaton, 2015; Marsh et al., 2020).

One of the most commonly studied frames of reference is the big-fish-little-pond effect (BFLPE; Marsh & Parker, 1984; Marsh & Basarkod, 2023). Recent research has focused on testing the

CONTACT Geetanjali Basarkod geetanjali.basarkod@acu.edu.au Institute for Positive Psychology and Education, Australian Catholic University, L 9, 33 Berry Street, North Sydney 2060

This article has been corrected with minor changes. These changes do not impact the academic content of the article.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10888438.2023.2174028.

universality of this phenomenon through studies using cross-national data (e.g., Marsh et al., 2020). Yet, much of this research has been dominated by studies on subjects with content that is relatively consistent across countries (e.g., math and science), while neglecting subjects whose content may differ (e.g., reading). In addition, research has shown self-concepts in the math and science domains to be vastly different – or even uncorrelated – to those in the reading domain (e.g., Möller & Marsh, 2013). Therefore, conclusions of universality drawn from the former may not necessarily extend to the latter. However, it is still important to consider because reading self-concepts are also important for educational outcomes (e.g., Sewasew & Koester, 2019). Therefore, the focus of the present investigation is to test the cross-national generalizability of the BFLPE in the domain of reading.

The big-fish-little-pond effect

Parents and teachers often believe in the benefits of educating children in selective schools. Selective schools are those where students are accepted based on their academic merit; they aim to cater for academically gifted students with high potential who may otherwise be without sufficient classmates of their own academic standard. It is, therefore, assumed that students attending such schools will have a brighter and more successful future. However, evidence tends to point to the contrary. In a seminal study conducted in the 1980s, the average ability level of a school negatively influenced individual students' academic self-concept, over and above the positive effects of individual-student ability on self-concept (Marsh & Parker, 1984).

This phenomenon occurs through a social comparison process (Festinger, 1954); humans evaluate their own abilities and form their self-perceptions by comparing themselves with others. People are motivated to make such comparisons to obtain plausible information for decision-making in daily life (e.g., Gruder, 1971). In terms of self-concept formation, students compare their own achievement in a specific domain with the achievements of their peers in the same domain (Marsh & Parker, 1984; Marsh & Seaton, 2015). That is, for the BFLPE, the peers are the frame of reference. Thus, for instance, a student who outranks their local peers in mathematics will have higher levels of math self-concept, while a student who performs below-average compared to their local peers in math will have lower levels of math self-concept, even if both students have the same level of math performance (Marsh & Basarkod, 2023).

In support of the social-comparison basis of the BFLPE, past research has shown that controlling for social comparison processes minimizes the BFLPE. Wang and Bergin (2017) showed this using data from 4th grade students in 59 countries/regions from the TIMSS 2011 dataset. Their results demonstrated that the negative effect of class-average math achievement on student math self-concept —which was statistically significant in 58 of the countries/regions examined—was eliminated in 6 and minimized in another 49 countries/regions, when accounting for perceived relative standing. Using a single-country sample from France, Huguet et al. (2009) explicitly designed a study to test this issue. They found that the negative effect of class-average achievement was eliminated after controlling for students' comparisons with their classmates. In a follow-up to the Huguet et al. (2009) study, Marsh et al. (2014) showed that accounting for an explicit measure of Dutch secondary school students' comparisons with their peers reduced the BFLPE for Dutch, English, and math. Thus, past research shows that social comparison processes do underlie the BFLPE, at least to some extent.

In the present study, we extend past research on the examination of the social comparison basis of the BFLPE by examining the effect of group-average achievement on three different reading self-belief measures. These are: reading self-concept of competence, reading self-concept of difficulty, and PISA test difficulty rating (a measure that we argue does not have the same social comparison basis as the reading self-concept measures). We describe these three measures—as well as our reasoning for the lack of social comparison processes underlying the third measure-in detail in the next section.

In recent BFLPE research, there has been an emphasis on replication, empirical tests, and cross-national generalizability. If evidence for a given theory is found in multiple countries and cultural contexts, it can be thought of as a universal principle with pan-human validity (Segall et al., 1998; also

see Marsh et al., 2020). There is now considerable empirical support for the cross-national generalizability of the BFLPE for multiple domains of self-concept, in both primary and high school students. For instance, Guo et al. (2018) found cross-cultural generalizability for the BFLPE for math, science, and reading in nationally representative samples of primary school students across 15 Organisation for Economic Co-operation and Development (OECD) countries. Their results revealed an average BFLPE beta estimate of −.32 for math, −.22 for reading, and −.16 for science. However, all 15 countries from this dataset were OECD countries, which are primarily developed countries (e.g., Australia, Canada, Germany). The lack of heterogeneity in the nature of countries examined limits our understanding of the extent to which the BFLPE is generalizable across different national contexts.

Studies using data from the Programme for International Students Assessment (PISA), in contrast, have provided cross-national generalizability for the BFLPE for high school students, across nationally representative samples from numerous developed and developing countries. Marsh et al. (2020) listed evidence for the BFLPE for general, math, and science domains across samples of 15-year-old students from 26–68 countries. They provided a summary of effect sizes for each country participating in the PISA rounds of 2000 (average $\beta = -.20$ for general self-concept in 26 countries; Marsh & Hau, 2003), 2003 ($\beta = -.31$ for math self-concept in 41 countries; Seaton et al., 2009), 2006 ($\beta = -.19$ for science self-concept in 57 countries; Nagengast & Marsh, 2012), and 2012 ($\beta = -35$ for math self-concept in 68 countries; Marsh et al., 2020). While these studies have provided substantial cross-national evidence for the BFLPE, they are limited to the domains of math and science.

The evidence briefly reviewed here highlights that there is no cross-national BFLPE study for reading which utilizes samples of high school students from a wide variety of developed and developing countries. Having such a study is crucial for two reasons. First, academic self-concept is multidimensional and hierarchical in nature, with the domains becoming increasingly differentiated with age (Marsh & Shavelson, 1985; Shavelson et al., 1976). Thus, as the only cross-national study (albeit only in OECD countries) of the BFLPE for reading self-concept was conducted using primary school students, the results cannot be generalized to older students. Second, in past cross-national studies, authors rationalize their focus on math and science subjects in terms of their content being relatively similar across countries (e.g., Marsh et al., 2020). Such a focus is at a disservice to other important achievement areas (e.g., reading) that may be less similar across countries. Indeed, self-concept in the reading domain is vastly different from those in the math and science domains, with verbal self-concepts being at the opposite end of the academic self-concept continuum as math self-concepts (with the various science self-concepts closer to math than to verbal self-concept).

Examining cross-national differences in reading is important because languages vary in many different ways. For instance, languages differ in terms of their orthographic depth, which is the reliability of print-to-speech correspondences. English—which has a deep orthography-is a much more difficult language to learn and to read compared to languages such as Finnish and Italian, which have shallower orthographies (Ziegler & Goswami, 2005). For instance, there are words in English that are spelled in the same way but are pronounced differently (heteronyms – tear [cry], tear [rip]). Likewise, there are words spelled differently that are pronounced the same (homophones – ate, eight). In contrast, languages like Finnish and Italian tend to have a more consistent one-to-one correspondence between how they are spelled and pronounced. A meta-analysis showed that because of English's relatively opaque orthography, early reading development can be different for children learning English than those learning languages with shallower orthographies (Florit & Cain, 2011). Past research has also shown these issues to contribute to differences in levels of reading comprehension in high school (Shankweiler et al., 1996). Thus, we cannot assume that results from more math-like subjects will translate to results from more verbal subjects.

Some past research has examined the effect of group-average achievement on reading in samples of students from single countries. Huguet et al. (2009) demonstrated that for secondary school students in France (ages 12–14, $N = 2{,}015$), class-average achievement negatively predicted French self-concept, while accounting for student-level achievement. Seaton et al. (2011) tested the BFLPE in reading using a nationally representative sample from Australia (age 15, $N = 5{,}176$). They found that school-average

achievement negatively and significantly predicted English self-concept. Thus, past research with single-country samples for reading self-concept shows a similar pattern to cross-national research in other domains regarding the negative effect of class-average achievement on self-concept. However, given that these studies are conducted on samples from developed countries, conclusions cannot be made about the generalizability of these results across countries.

Therefore, we extend the examination of the cross-national generalizability of the BFLPE to the domain of reading for high school students, using data from the PISA2018 cohort involving nationally representative samples from 72 countries. We test the BFLPE for two subcomponents of reading self-concept (competence and difficulty), as well as for ratings of difficulty experienced in solving items on the PISA reading test. We describe all three reading self-beliefs in the next section.

Reading self-beliefs

Self-concept (competence and difficulty)

Self-concept of perceived competence assesses the feelings of proficiency and skill in reading and reading-related activities (e.g., "I am a good reader"). Self-concept of perceived difficulty measures the extent to which individuals have difficulty with, or aversion to, reading and reading-related activities, events, and behaviors (e.g., "I have always had difficulty with reading;" see Measures section for all items used in PISA2018). Thus, while perceived competence assesses a positively-valenced self-belief, perceived difficulty assesses a negatively-valenced self-belief. The assessment of both is designed to contribute to a more complete understanding of an individual's reading self-concept (Chapman & Tunmer, 1995).

PISA test difficulty rating

The PISA2018 cohort also included a measure of the level of difficulty experienced by the students for the reading tasks within the PISA test assessment (see Measures section for items). We refer to this construct as the *PISA test difficulty rating*, to differentiate it from the perceived difficulty subscale of reading self-concept. Although both measures assess difficulty in reading, PISA test difficulty rating is concerned specifically with the PISA reading assessment, while self-concept of perceived difficulty concerns reading in general.

Absolute versus relative frames of reference

We note that neither the self-concept items nor the PISA test difficulty rating items explicitly asked students to compare their reading abilities with their peers. The critical feature of the PISA test difficulty rating items, therefore, is that the self-evaluations a student needed to undergo to answer these questions was based entirely on the specific task at hand (the PISA reading assessment). Social comparisons are more likely when an individual is uncertain of their standing (Brown et al., 2007). This implies that evaluations based on specific criteria provide less of an opportunity for comparisons with others (Marsh, Pekrun, et al., 2019). Therefore, because of this well-defined context, the PISA test difficulty rating has an absolute frame of reference, minimizing invocation of social comparison processes. In contrast, as the items for self-concept of perceived difficulty are about reading in general, there is more of an opportunity to compare one's reading difficulties with those of peers. Therefore, it can be said to have a more relative frame of reference.

Previous research has juxtaposed self-belief measures in relation to this absolute vs relative frame of reference difference. The most salient differentiation is between self-concept and self-efficacy (i.e., belief about one's capabilities to produce designated levels of performance; Bandura, 1994). While there are several similarities between the two constructs (e.g., perceived competence, domain specificity), social comparison processes are central to the distinction between them (Bong & Skaalvik, 2003). Self-concept ratings invoke social comparison processes with evaluations of self-relative-to-others. In

contrast, self-efficacy measures tend to solicit goal-referenced evaluations that are task-specific and in-built (e.g., "how confident do you feel about having to solve an equation like $3 \times +5 = 17$;" OECD, 2014). Thus, self-efficacy measures are not dependent on comparisons with others. Marsh, Pekrun, et al. (2019) note that many so-called self-efficacy measures are really more like self-concept measures in that they invoke social comparison processes to respond to the items, by not including the goal reference within the items. Hence, they demonstrated that BFLPEs were substantially minimized for pure self-efficacy measures based on absolute standards of comparison but were similar to BFLPEs based on self-concept ratings for so-called self-efficacy measures based on relative standards of comparison.

In the present investigation, we seek an alternative test of the social comparison process underlying the BFLPE based on absolute vs. relative frames of reference. The PISA test difficulty rating is fundamentally different from pure self-efficacy ratings in that self-efficacy ratings are prospective in relation to future behavior, whereas the PISA test difficulty ratings are retrospective in relation to the PISA test that students had already taken. Nevertheless, pure self-efficacy and PISA test difficulty ratings share a reliance on an absolute standard of comparison that are task-specific and do not require social comparison processes to form a rating. Thus, we further extend past research to evaluate the effect of school-average achievement on student reading beliefs across countries, not only in relation to the two dimensions of reading self-concept that assess domain-specific self-beliefs that allow for social comparisons, but also a measure specific to the PISA reading test that assesses task-specific beliefs with an absolute frame of reference.

The present study

The BFLPE is one of the most well-supported findings in social psychology (Marsh & Seaton, 2015). However, much of the evidence for its cross-national generalizability has been provided in the domains of math and science. There have been no cross-national studies of the BFLPE for reading in samples of secondary school students, nor any that have explored the BFLPE separately for reading self-concept of competence and difficulty. The present study aims to address these three issues by using data from the PISA2018 cohort, which comprises responses from over half a million 15-year-old students from 72 countries/regions. We explore the effect of school-average achievement on reading self-concept of competence and difficulty, and draw a parallel with its effect on PISA test difficulty rating. We explore these effects first across the whole sample, and then for each country.

A recent cross-national study of 10-year-old students found that the association between achievement and self-concept was slightly stronger for difficulty than for competence (Lafontaine et al., 2019). However, Chapman and Tunmer (1995) stated that responses on the two self-concept scales become more similar with an increase in age, as do their correlations with achievement. Although previous research on the association between achievement and the two self-concept scales has only been considered at the individual level, we extrapolate to predict the following:

Hypothesis 1: There will be BFLPEs for both reading self-concept of perceived competence and perceived difficulty (**H1a**), and these effect sizes will be comparable to each other **(H1b)**.

Previous literature shows that the BFLPE is based on social comparisons (Marsh & Seaton, 2015). Items assessing reading self-concept of competence and difficulty allow for such social comparisons, while those assessing PISA test difficulty do not. Following from the Marsh, Pekrun, et al. (2019) study that showed the BFLPE to be substantially truncated for self-efficacy ratings that have a similar absolute frame of reference, we hypothesize the following:.

Hypothesis 2: There will not be a BFLPE for the PISA test difficulty rating.

Past studies on the cross-national generalizability of the BFLPE have either used samples of primary school students in OECD countries to examine the domain of reading (Guo et al., 2018), or of high school students in a broad range of countries to examine the domains of math and science (Marsh et al., 2020). As the BFLPE in these studies have been fairly consistent across the countries examined, we extend past findings to predict the following for reading self-beliefs in high school students:

Hypothesis 3: The effect of school-average achievement on reading self-concept of competence will be negative in all countries (**H3a**), and positive in all countries for reading self-concept of difficulty (i.e., the opposite direction of the effect of student-level achievement; **H3b**).

Hypothesis 4: The effect of school-average achievement on PISA test difficulty rating will be weak or non-significant in all countries.

Methods

Data and participants

We used the cross-national PISA2018 dataset, which is publicly available through the OECD website (oecd.org/pisa/pisaproducts/). Students completed tests to assess their knowledge and skills in reading, math, and science, and answered questionnaires assessing a variety of psychosocial and family background variables. PISA utilizes a two-stage stratified sampling design. The first stage systematically samples schools from a comprehensive national list of all eligible schools with 15-year-old students. The second stage involves sampling thirty-five (where available) 15-year-old students from each school. These data constitute nationally representative samples after using the appropriate survey weights provided as part of the public database (OECD, 2019). The authors' university does not require ethics approval for secondary data analyses of publicly available data such as PISA.

The raw PISA2018 dataset consists of responses from 612,004 students from 21,903 schools in 80 countries/economic regions. For the present study, data from countries with 100% missingness on our key variables were removed. Specifically, data from Spain and Vietnam were removed because these countries did not have reading achievement scores, data from Lebanon and Macedonia were removed because they did not answer the reading self-concept questions, and data from Belarus, Russia (mainland), Morocco, and Malaysia were removed as the OECD reported that the composite scores for reading self-concept of competence and difficulty did not perform well in these countries (with the PISA test difficulty rating also not performing well in Morocco) and, therefore, did not provide these in the dataset. Thus, analyses reported in this paper used data from 533,165 students from 19,366 schools in 72 countries. The remaining small amount of missing data was dealt with using multiple imputations (see Statistical Analyses section).

Measures

The PISA survey consists of achievement tests and student background questionnaires. The test component lasts for approximately 2 hours and assesses students' proficiency in reading, mathematics, and science. The questionnaire component lasts for approximately 30 minutes and involves answering a series of questions on background information and psycho-social constructs. The entire survey testing experience for a student usually lasts 3–3.5 hours, including instructions and break periods. Information about all variables used in the present study, including overall and country-by-country psychometric properties, is available from the PISA technical report (oecd.org/pisa/data/pisa2018technicalreport/). Information about means, standard deviations, and percent of missing data before

imputations for all student-level variables and covariates are presented in the Supplementary Materials (Section 1, Table S1).

Dependent variables

We utilize the total scores (Warm estimates) for all our dependent variables, that were calculated by PISA and provided in the dataset (OECD, 2019).

Reading self-concept (competence and difficulty). Reading self-concept was measured using two subscales; perceived competence and difficulty. Three items assessed perceived competence ("I am a good reader," "I am able to understand difficult texts," and "I read fluently"), and the other three assessed perceived difficulty ("I have always had difficulty with reading," "I have to read a text several times before completely understanding it," and "I find it difficult to answer questions about a text"). All items were rated on a scale of 1 (*strongly disagree*) to 4 (*strongly agree*). Higher scores on the reading competence subscale are indicative of greater perceived competence in reading, while higher scores on the difficulty subscale are indicative of greater perceived difficulty with reading.

PISA test difficulty rating. Following a testing break, students were asked about the level of difficulty they experienced specifically in the PISA reading test. This was measured using three items with the stem "In the PISA test you took before the break, you had to read several texts and answer reading comprehension questions. How do you feel about these reading tasks?" Each item (i.e., "There were many words I could not understand," "Many texts were too difficult for me," and "I was lost when I had to navigate between different pages") was rated on a scale of 1 (*strongly disagree*) to 4 (*strongly agree*). The PISA test difficulty rating is a unidimensional measure, and the OECD reports that the measure had adequate reliability in all countries, except for Morocco (data from which was deleted from our study; please see the section "Data and Participants" above). The average reliability of this measure for the countries included in our analyses was 0.815 (range = .710 to .888; calculations based on data presented in Tables 16.23 and 16.24 in the PISA 2018 Technical Report).

Independent variables

Student-level reading achievement. PISA's measure of reading achievement represents the reading literacy exhibited by the test-taker in terms of their ability to understand, use, evaluate, reflect on, and engage with texts. It included scenarios in which students were provided an overarching purpose for reading a collection of thematically related texts to complete a higher-level task. The responses were in a combination of multiple choice and free text formats. For instance, students are asked to read a blog post written by a professor on her trip to Rapa Nui (Easter Island). Then, students are asked seven questions about this blog post. An example of a multiple-choice question is: "According to the blog, when did the professor start her field work." To answer this question, students needed to choose from: (a) During the 1990s; (b) Nine months ago; (c) One year ago; or (d) At the beginning of May. An example of a free-text response question is: "In the last paragraph of the blog, the professor writes: 'Another mystery remained . . . ' To what mystery does she refer?"

To reduce the testing burden on students, each student responded to a subset of items that measure their reading achievement. Given this missing by design procedure, PISA provided 10 plausible values (PVs; see Statistical Analyses section) for each student's reading achievement to allow researchers to account for this uncertainty in their analyses. These PVs were obtained via models using item response theory to infer students' abilities based on their performance on the PISA test (see the PISA 2018 data manual for details). Higher scores are indicative of greater reading literacy/achievement. We note here that PISA undergoes extensive testing to establish measurement equivalence across countries and language of assessment, for all their achievement items. Please see Supplementary Materials (Section 2) for more information about the reading achievement measure as well as the equivalence testing procedure.

Table 1. Pearson's correlations for all student-level study variables.

		1	2	3	4	5	6	7	8	9	10
1	Reading Competence										
2	Reading Difficulty	−.319									
3	PISA Test Difficulty	−.310	.493								
4	Student-Level Reading Achievement	.259	−.264	−.370							
5	Age	.013	−.004	−.015	.033						
6	Gender	−.078	.027	−.006	−.138	*.001*					
7	SES	.186	−.153	−.257	.387	−.007	.011				
8	Modal Grade	.084	−.064	−.102	.242	.255	−.057	.210			
9	Grade Repetition	−.102	.087	.114	−.254	.008	.061	−.183	−.447		
10	Immigrant Status	.030	.007	.007	.014	.011	.008	.056	.009	.059	
11	Language Spoken at Home	−.025	.062	.080	−.121	*.000*	.013	−.036	.024	.060	.303

Correlations in italics are not statistically significant (i.e., $p > .05$). Gender: 1 = females and 2 = males; grade repetition: 0 = no, 1 = yes; immigrant status: 1 = natives, 2 = second-generation immigrants, 3 = first-generation immigrants; language spoken at home: 0 = language of testing and 1 = not language of testing.

School-average reading achievement. First, each student-level reading achievement PV was grand-mean centered. Then, we calculated school-average reading achievement as the average student reading achievement in each school, independently for each PV. The resulting 10 sets of values for school-average reading achievement were used in the same way as the PVs for student-level reading achievement (see Statistical Analyses section).

Covariates

We controlled for several student-level covariates typically used in BFLPE studies (e.g., Marsh et al., 2020). Specifically, we accounted for students' gender, age, SES (a composite index of economic, social, and cultural status), year in school (grade compared to modal grade in the country), grade repetition, immigrant status, and language spoken at home (see Table 1 notes for scoring of variables). All covariates were grand-mean centered. Table 1 presents correlations amongst all student-level variables and covariates.

Statistical analyses

We conducted all analyses in R (Version 3.6.2; R Core Team, 2019), the code for which is available through the Open Science Framework (https://osf.io/m296y/). The final data were multiply imputed using the Amelia II package (Honaker et al., 2011). We created ten imputed datasets, retaining all variables and covariates mentioned above. We assigned each student-level and school-average reading achievement PV to one of the 10 imputed datasets, in accordance with the PISA data manual.

We then conducted all analyses separately in each dataset and then combined them using Rubin's (1987) rules. Correlational analyses were conducted using the miceadds package (Robitzsch & Grund, 2019) and multilevel models using the lme4 package (Bates et al., 2015). For the multilevel models, we included random intercepts at the school and country level to account for clustering of students within schools, and of schools within countries. We included fixed effects of student-level and school-average reading achievement scores as well as their random effects at the country level. All models were weighted using the PISA-provided final survey weight, normalized for each country (i.e., sum of the weights in each country was equal to the sample size in that country). We present models with covariates in the main manuscript, although results were similar when we did not include covariates (Supplementary Materials: Section 3, Table S2).

We interpret the size of the fixed effects based on recommendations by Else-Quest et al. (2010) who stated that an effect size of less than .1 can be considered trivial when sample sizes are large, as estimates of even .01 may be statistically significant. We interpret random effects, and the generalizability of fixed effects across countries, based on recommendations from Marsh (2016). Marsh suggested that there is good support for the generalizability of an a-priori prediction if the standard

deviation of country-to-country variation (i.e., the square root of the random variance component) is less than half that of the corresponding fixed-effect estimate in support of that prediction (as the direction of the effect will not change even at relatively extreme values).

We also graphed forest plots containing beta estimates (with 95% confidence intervals) for the effect of school-average achievement on our outcome variables in each country, to further explore country-to-country variations in these effects. These beta estimates were extracted from two-level models (clustered at the school level) run independently in each country, and included all covariates.

Results

Multilevel models

Fixed effects

Table 2 displays results from three survey-weighted multilevel models (our main cross-national analyses). Results indicate that the main effect of student-level reading achievement was positive and significant for self-concept of reading competence, and negative and significant for self-concept of reading difficulty and PISA test difficulty rating. The magnitude of this effect was similar for all three dependent variables ($|\beta| = .317–.348$).

The main effect of school-average reading achievement, our test of the BFLPE, was in the opposite direction to that of student-level reading achievement for all models. Thus, the higher the school-average achievement, the lower the student's reading self-concept of perceived competence, perceived difficulty, and PISA test difficulty rating (controlling for individual-level achievement and our covariates). However, the magnitude of the effect of school-average achievement varied considerably across the different reading self-belief measures. Specifically, reading competence showed the strongest BFLPE ($\beta = -0.198$), followed by reading difficulty ($\beta = 0.119$). While the effect of school-average reading achievement for PISA test difficulty rating was significant, the effect size was close to zero ($\beta = 0.028$). Thus, the effect of school-average achievement on the PISA test difficulty rating may be considered trivial in size (Else-Quest et al., 2010). Similar results for all three reading self-beliefs were found when utilizing the five subcomponents of the PISA reading achievement measure (for an explanation of these subcomponents and their corresponding results, please see Supplementary Materials: Section 4, Tables S3–7). Our findings provide support for Hypothesis 1a (BFLPEs were present for the two reading self-concept subscales), do not support Hypothesis 1b (the BFLPE was larger for reading self-concept of perceived competence than for reading self-concept of perceived difficulty), and provide partial support for Hypothesis 2 (weak, but negligible BFLPE for PISA test difficulty rating).

Random effects

Following recommendations from Marsh (2016), although the country-to-country variation for the effect of school-average achievement was largest for self-concept of perceived competence than for the other two outcomes based on random effects alone ($SD = 0.087$), the BFLPE for competence was the most generalizable across countries given the large main effect of school-average achievement ($\beta = -0.198$; this beta estimate is larger than two times the random effect standard deviation). Thus, it can be said that the negative effect of school-average achievement on self-concept of perceived competence has cross-national generalizability. The generalizability of the effect of school-average achievement across countries was marginally weaker for self-concept of difficulty and much weaker for PISA test difficulty rating, given that their SDs were more than half of the fixed-effect (self-concept of perceived difficulty: $SD = 0.071$; $\beta = 0.119$; PISA test difficulty rating: $SD = 0.067$, $\beta = 0.028$).

Table 2. Results from survey-weighted multilevel models predicting reading self-concept subscales and PISA test difficulty rating.

	M1: SC Competence		M2: SC Difficulty		M3: PISA Test Difficulty	
	β	*SE*	β	*SE*	β	*SE*
Fixed Effects						
Intercept	−.028	.031	−.034	.024	.002	.031
Student-Lvl Reading Ach	.348*	.010	−.317*	.008	−.340*	.010
School-Avg Reading Ach	−.198*	.011	.119*	.010	.028*	.011
Covariates						
Age	.006*	.001	.000	.002	−.004*	.001
Gender	−.041*	.001	−.008*	.001	−.044*	.001
SES	.130*	.002	−.073*	.002	−.110*	.002
Modal Grade	−.016*	.002	.013*	.002	.001	.002
Repeated Year	−.021*	.002	.010*	.002	.011*	.002
Immigration Status	.020*	.002	−.005*	.002	.004*	.002
Language Spoken at Home	−.018*	.002	.016*	.002	.025*	.002
Random Effects	*SD*		*SD*		*SD*	
School-Lvl Intercept	.148		.123		.130	
Country-Lvl Intercept	.266		.206		.204	
Student-Lvl Reading Ach| Country	.084		.066		.069	
School-Avg Reading Ach | Country	.087		.071		.067	
Residual	.901		.931		.888	

Parameter estimates (β) for fixed effects that differ from zero by more than two standard errors (*SE*) are statistically significant ($p < .05$), and are indicated by *. Random effects are presented in standard deviations (*SD*).

Country-specific BFLPEs

To get a better sense of the country-to-country variation in the effect of school-average achievement on the three dependent variables, we present forest plots (Figure 1). These forest plots graphically represent the country-to-country variation in the beta estimates for school-average reading achievement for each of our three dependent variables (we provide the exact numerical values for effect sizes and standard errors for each country in Supplementary Materials: Section 5, Table S8). In terms of direction, the effect of school-average achievement was most consistent across countries for reading self-concept of perceived competence, and least consistent for PISA test difficulty rating.

Specifically, the effect of school-average reading achievement on student-level reading self-concept of perceived competence was negative and significant in 66 of the 72 countries (β = −0.070 to −0.369), and never positive. Thus, it suggests that while the magnitude of the BFLPE for reading competence showed some country-to-country variation, it was negative in all countries, and significantly so in most (supporting Hypothesis 3a). The effect of school-average reading achievement on reading self-concept of perceived difficulty was also consistent; all the significant effects were positive (56 countries, β = 0.053 to 0.292), although the non-significant effects were on both sides of zero (positive in 12 countries, negative in 4 countries; partial support for Hypothesis 3b). The effect of school-average reading achievement on student-level PISA test difficulty rating was non-significant in most countries (45 countries; partial support for Hypothesis 4), and even the significant effects were on both sides of zero (19 positive and 8 negative; β = −0.111 to 0.201). Thus, our results provide support for the cross-national generalizability of the BFLPE for self-concept of competence and difficulty, as well as for the weak effect of school-average achievement on the PISA test difficulty rating.

Discussion

Our study is the first to: (i) examine the BFLPE for reading in secondary school students cross-nationally, (ii) test the generalizability of the BFLPE for reading self-concept over a wide variety of countries (i.e., both developing and developed countries) and (iii) test the BFLPE across two separate subscales of reading self-concept as well as for difficulty experienced specifically during the PISA reading test. Results indicated that the BFLPE for reading self-concept of 15-year-old students was robust across the 72 countries/regions examined (particularly for perceived competence). In addition,

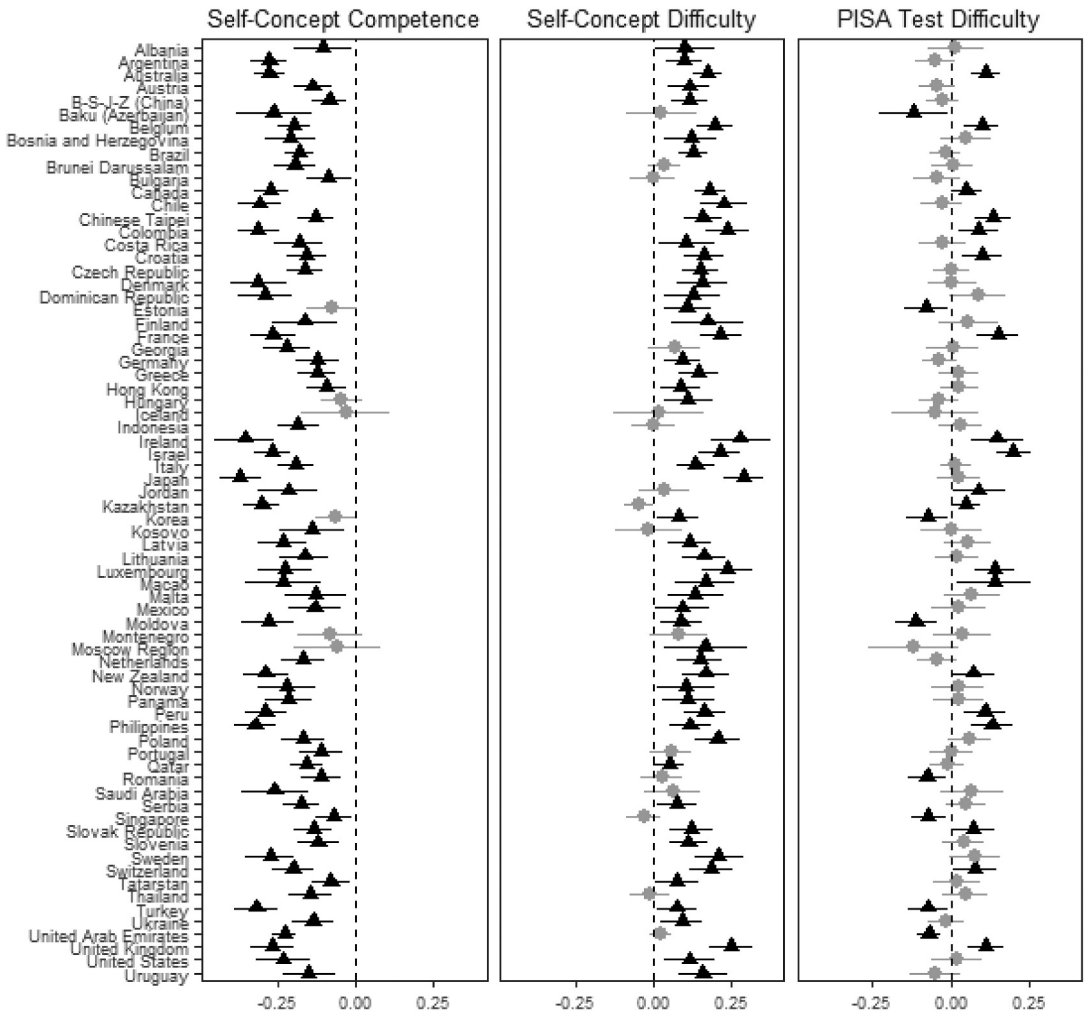

Figure 1. Country-by-country variation in beta estimates (with 95% confidence intervals) for school-average achievement predicting reading self-beliefs.
NOTE: Beta estimates are taken from 2-level models run separately in each country, for each dependent variable, controlling for student-level reading achievement and covariates. Black triangles indicates significant estimates (95% CI line does not cross zero), while gray circles indicates non-significant resultsestimates (95% CI line crosses 0).

the effect of school-average achievement on individual-level experienced difficulty with the PISA reading test (PISA test difficulty rating) was non-significant in most countries, showing that the social comparison process was largely not in play for this construct.

Social comparison processes

Previous research has provided evidence for social comparison process as the mechanism underlying the BFLPE. That is, in order to evaluate their performance, students compare their abilities in a certain domain with the abilities of their peers in that domain. We provide a test of this underlying mechanism by contrasting the effects of school-average achievement on measures that allowed for such social comparisons by keeping the self-evaluations general to a domain (self-concept of competence and difficulty), with those on measures that specified the basis of such evaluations to a defined task (PISA test difficult rating).

Differences between the two self-concept subscales

There was substantial support for the BFLPE—and the cross-national generalizability of the BFLPE—for both self-concept subscales. However, it is interesting to note that the associations of these two subscales were of a more similar magnitude with student-level reading achievement ($|\beta| = 0.348$ and 0.317, respectively) than with school-average reading achievement ($|\beta| = 0.198$ and 0.119, respectively).

Why is the BFLPE larger for self-concept of perceived competence than perceived difficulty? Such a pattern may imply that social comparison processes operate more strongly for evaluations of self-concept of competence than for self-concept of difficulty. This might be due to the valence of the items assessing the two subscales. Self-concept of competence measures a positively-valenced self-evaluation (e.g., I am a good reader), but self-concept of difficulty measures a negatively-valenced self-evaluation (e.g., I have always had difficulty with reading). Fredrickson (2001) broaden-and-build theory posits that positive emotions increase an individual's levels of attention and awareness, and serve as instruments of individual growth and social connections, whereas negative emotions do so to a lesser extent. While self-concept is considered a motivation rather than an emotion (Basarkod & Marsh, 2023), it is possible that social comparisons regarding differently valenced evaluations of self-concept follow the same broaden-and-build pattern. Evaluations of competence may lead individuals to be more open, enabling them to acknowledge others' levels of competence and their own standing in relation to them. In contrast, evaluations of difficulty may cause individuals to shut off and be less aware of their peers' levels of difficulty and their standing in relation to these peers, thereby reducing the strength of social comparisons. Future research is, however, required to confirm this supposition.

Differences between self-concept perceived difficulty and PISA test difficulty

Our results demonstrate that the BFLPE is severely truncated when there is an absolute frame of reference for evaluations of reading self-beliefs. That is, while we find a BFLPE for the self-concept perceived difficulty items that were generic to the domain of reading, this effect was negligible when students were provided with an absolute frame
of reference on which to base their evaluations of reading difficulty. Although neither the items assessing self-concept of perceived difficulty nor those measuring PISA test difficulty ratings explicitly asked students to rate their difficulty with reading in comparison to their peers, it is apparent that the social comparison process operated more strongly for the former than for the latter.

These findings contrasting relative and absolute frames of reference are in line with past research that show the BFLPE to be stronger for self-concept than for self-efficacy (Marsh, Pekrun, et al., 2019). As mentioned in the Introduction, items assessing what Marsh, Pekrun, et al. (2019) referred to as pure self-efficacy also have an absolute frame
of reference, in that they explicitly state the criterion against which students need to evaluate themselves. However, while PISA test difficulty was a *retrospective* rating with an absolute frame of reference (i.e., evaluations based on a specific task students already completed), self-efficacy items involve a *prospective* rating with an absolute frame
of reference (i.e., evaluations based on a specific task students imagine themselves doing in the future). In addition, past research has only compared self-concept and self-efficacy in relation to the BFLPE using scales assessing positive evaluations (e.g., of competence), or a mix of positive and negative evaluations, (e.g., mix of items assessing both competence and difficulty), but not both separately (Marsh, Pekrun, et al., 2019). Despite these differences, our results were matched with those from past studies comparing measures of self-concept and self-efficacy. Still, future research is needed to explicitly juxtapose the effect of school-average achievement on measures of self-concept with its effect on absolute frame
of reference measures that are prospective vs retrospective, and negatively- vs positively-valenced.

One avenue for future research is to examine the predictive utility of PISA test difficulty ratings in predicting later achievement, while controlling for prior achievement. This would allow researchers to test how closely the PISA test difficulty rating is aligned with achievement. If PISA test difficulty does

not uniquely predict later achievement while controlling for previous achievement, we can conclude that the two are closely aligned. As our study utilized cross-sectional data, future longitudinal research is required to test this. Such research could also examine whether the PISA test difficulty rating performs differently to measures of self-concept in predicting future achievement.

We do note here that the correlation between achievement and PISA test difficulty rating ($r = -.370$) indicates that there is a general trend for students who perform higher to rate their perceived difficulty as being lower. However, its magnitude implies that there is still room for factors other than student-level achievement to impact students' perceptions of their difficulty experienced during the PISA test. These factors could be phenomena such as the positive illusory bias and the self-protective hypothesis, where for instance, students may report the test was less difficult than they actually found it (e.g., Heath & Glen, 2005; Hoza et al., 2002). In addition, our multilevel models demonstrated that demographic factors might also impact students' ratings of PISA test difficulty. Specifically, given the same level of student- and school-level achievement, female students, students from low SES backgrounds, students who repeated a year at school, and those who did not speak the language of testing at home had higher ratings of difficulty experienced during the PISA reading test. As this is the first time that PISA asked questions about the level of difficulty students experienced during the PISA test, future research is required to more fully understand this measure and the way in which students respond to it.

Cross-national generalizability of results

In the universalist perspective in cross-cultural psychology, there is an emphasis on replicability of results, empirical tests, and cross-cultural generalizability of support for theoretical predictions. Such examinations of the generalizability of a given effect are traditionally conducted using meta-analyses. However, the extent to which meta-analyses can provide evidence for the cross-national generalizability of an effect depends substantially on the degree to which the individual studies contained within it represent samples from varied countries. Indeed, the only known meta-analysis on the BFLPE to date (Fang et al., 2018), while consisting of studies using varied samples for math and science subjects, consisted largely of Western, educated, industrialized, rich and democratic (WEIRD) samples for reading self-concept (with 10 effect sizes from Europe and America, and two from Asia). The use of large-scale cross-national datasets, such as the PISA data used here, can help overcome such drawbacks of meta-analysis (see Marsh et al., 2020 for a detailed discussion on the juxtaposition of meta-analytic studies and cross-national studies based on PISA data).

The most salient finding of the current investigation was that the BFLPE for reading self-concept of competence was never positive, and the BFLPE for reading self-concept of difficulty was never significantly negative. The remarkable cross-national support for the BFLPE for reading self-concept suggests that the mechanism upon which this effect is based (i.e., social comparison) constitutes a cross-nationally valid theory of self-concept formation. While previous studies have provided such evidence in relation to general, math, and science self-concepts, the current study is the first to provide robust cross-national support for the BFLPE for reading self-concept. This finding is important in that it highlights that social comparison processes are universal, even when the content of the subject under consideration may vary between countries (i.e., different countries speak different languages, and different languages have different rules in terms of grammar, syntax, etc).

While there was robust support for the cross-national generalizability in the direction of the BFLPE—especially for the reading self-concept of competence subscale—there were differences in the magnitude of these effects across countries. Though studies have shown the BFLPE to be robust to moderators at the student-level (e.g., Marsh et al., 2020), there is little past research examining country-level moderators. However, some past research has suggested that BFLPEs for math self-concept tend to be stronger in countries with greater levels of ability stratification (i.e., more selective schooling; Parker et al., 2021; Salchegger, 2016). To explain the differences in BFLPEs in our study post-hoc, we examined whether ability stratification (measured via intra-class correlations of reading

achievement; Parker et al., 2021) the Human Development Index (HDI; a standard measure of country-level development), the Gini Index (a standard measure of country-level equality), or country-average achievement were correlated with the BFLPEs for the self-concept of competence and difficulty subscales. Surprisingly, ICCs were not correlated with either subscale. This further highlights that reading – and the BFLPE for reading – may be different to other subjects such as math (for which ICCs have been shown to moderate effects; P. Parker et al., 2021). Similarly, neither HDI nor Gini were correlated with BFLPEs for either of the self-concept scales. In contrast, country-average achievement had a positive correlation with the BFLPE for self-concept of difficulty ($r = .39$). This indicates that the effect of school-average achievement on students' self-concept perceptions of difficulty was stronger in countries with greater levels of country-average achievement.

However, this result should be interpreted with caution as the correlation between ICC and BFLPE was only significant for the reading self-concept of difficulty subscale, and not the competence subscale (which much of the past research on reading self-concept has utilized). A recent study by Basarkod et al. (2022) might provide insight into these results. They examined the dimensionality of reading self-concept using the same PISA 2018 data presented here across levels of student reading achievement. Their results demonstrated that while students with high levels of achievement gave consistent responses to the two subscales (i.e., high levels of perceived competence corresponded with low levels of perceived difficulty, and vice versa), students with lower levels of achievement gave inconsistent responses to the two subscales (i.e., students simultaneously reported high levels of perceived competence and high levels of perceived difficulty, and vice versa). These results seemed to be largely due to responses on the difficulty subscale, with students with low levels of reading achievement having trouble responding to these items. Indeed, past research has shown that individuals have trouble responding to questions that are negatively valenced, such as the self-concept of difficulty items utilized here (e.g., Borgers et al., 2004). It is possible that because of the issues in responding to the self-concept of difficulty subscale for students with lower levels of achievement, the BFLPE for the reading self-concept of difficulty subscale was not as strong for countries with low levels of country-average achievement. However, future research is required to confirm this speculation and assess in detail the reasons why some countries show bigger BFLPEs than others, and why some have non-significant effects. Such research would ideally be conducted across several subject domains, to ascertain whether and why some country-level variables (e.g., ability tracking) affect BFLPEs in some domains and not others.

In terms of comparisons of the magnitude of the BFLPE with previous cross-national studies with math and reading (Marsh et al., 2020), the effect for self-concept of competence ($\beta = -0.198$) and difficulty ($\beta = 0.119$) were smaller than that for the BFLPE for math self-concept in PISA2003 ($\beta = -0.31$) and 2012 ($\beta = -0.35$). The BFLPE for competence was similar to that for science self-concept in PISA2009 ($\beta = -0.19$), while the BFLPE for difficulty was smaller. This is partly in line with the only known cross-national study that examined, in parallel, the size of the BFLPE for math, science, and reading (Guo et al., 2018), which showed that the BFLPE for reading ($\beta = -0.22$) was smaller than that for math ($\beta = -0.32$), but larger than that for science ($\beta = -0.16$). Interestingly, although the Guo et al. (2018) study was conducted using data from primary school students from 15 OECD countries, the magnitude of the average BFLPE for reading self-concept in their study, was comparable to that for reading self-concept of competence in our study of high school students from 72 developing and developed countries. While BFLPEs are expected to be larger for older students, this trend may be masked when examining a broader range of developed and developing countries. Future research that examines the BFLPE for primary school students using samples from both developed and developing countries can shed more light on this issue.

Limitations and directions for future research

Our findings, and our inferences based on them, require caveating. The present investigation is based on a single wave of cross-sectional data, limiting our inference regarding causality. However, various

combinations of longitudinal, quasi-experimental, matching and true experimental designs, and even introspective diary studies have consistently provided support for our underlying assumption that school-average achievement negatively predicts individual level self-concept (although usually for the math domain; e.g., Alicke et al., 2010). In addition, although we compare results based on self-concept measures and the PISA test difficulty measure, a stronger test of the relative vs. absolute frame of reference would be to explicitly state the standard of comparison against peers for the self-concept items, and making the absolute nature of the PISA test difficulty rating more explicit (e.g., asking students what percentage of the test items they think they answered correctly). Still, we provide initial evidence for the generalizability of the difference in effect of school-average achievement between items that did and did not have an absolute frame of reference, extending such results beyond measures of self-efficacy used previously (e.g., Marsh, Pekrun, et al., 2019).

We also note that the PISA measure of reading achievement is a general, multi-faceted measure of reading literacy. This measure is, therefore, likely assessing several sources of individual differences in reading literacy (e.g., fluency, vocabulary, comprehension). While we demonstrated similar results when using the subcomponents that were included in the PISA database (e.g., locating information, understanding and reflecting), these subcomponents do not capture all aspects of reading literacy (e.g., fluency). It is possible that these other aspects of reading have differential links with self-concept. For instance, the ability of a student to understand words (e.g., vocabulary or reading comprehension) may have a larger effect on their self-concept than would reading fluency. If this is the case, it would also be possible for school-average ability in understanding words to have a stronger negative effect on a student's self-concept, as compared to the effect of school-average reading fluency. While our subcomponent analyses are a step in the right direction to understand these differences, future research is required to assess the numerous components of reading independently and juxtapose the effects with each other. That said, given the consistency of results across our main analyses and sensitivity analyses, we would expect the links between the components of reading achievement (and their group-average counterparts) and self-concept to vary in the size of their estimates, rather than in their direction. That is, we suspect that greater performance in any component of reading would be linked with greater reading self-concept, and greater school-average reading performance to be linked with lower reading self-concept, but that the strength of these effects may differ from component to component.

Conclusion

Our study addresses several gaps in the BFLPE literature and provides robust cross-national evidence for the BFLPE for reading self-concept. Our findings of the universality of this effect are important considering the close association between self-concept and other important educational outcomes including achievement and long-term aspirations. Given the reciprocal relationship between achievement and self-concept (Marsh & Seaton, 2015), a diminished reading self-concept due to comparisons with peers is likely to lead to diminished reading achievement (Marsh, 1987), consequently preventing students from reaching their full potential in this domain (e.g., Marsh et al., 2007). Parents and teachers should take this into consideration when contemplating the consequences of high-achieving /selective schools on the educational outcomes of students.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Geetanjali Basarkod http://orcid.org/0000-0003-3708-2751
Herbert W. Marsh http://orcid.org/0000-0002-1078-9717
Jiesi Guo http://orcid.org/0000-0003-2102-803X
Theresa Dicke http://orcid.org/0000-0001-8868-2242
Kate Xu http://orcid.org/0000-0002-1585-1938
Philip D. Parker http://orcid.org/0000-0002-4604-8566

References

Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science*, *21*(2), 174–177. Academic Press.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior 4*, 71–81. Academic Press.

Basarkod, G., & Marsh, H. W. (2023, in press). Academic self-concept: A central motivational construct in educational psychology. In M. Bong, S. Kim, & J. Reeve (Eds.), *Motivation science: Controversies and insights*. (pp. 59–64). Oxford University Press. https://doi.org/10.1093/oso/9780197662359.003.0010

Basarkod, G., Marsh, H. W., Sahdra, B. K., Parker, P. D., Guo, J., Dicke, T., & Lüdtke, O. (2022). The dimensionality of reading self-concept: Examining its stability using local structural equation models. *Assessment*. https://doi.org/10.1177/10731911211069675

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Borgers, N., Sikkel, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity*, *38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Brown, D. J., Ferris, D. L., Heller, D., & Keeping, L. M. (2007). Antecedents and consequences of the frequency of upward and downward social comparisons at work. *Organizational Behavior and Human Decision Processes*, *102*(1), 59–75. https://doi.org/10.1016/j.obhdp.2006.10.003

Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, *87*(1), 154. https://doi.org/10.1037/0022-0663.87.1.154

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103. https://doi.org/10.1037/a0018053

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology*, *9*, 1569. https://doi.org/10.3389/fpsyg.2018.01569

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, *23*(4), 553–576. https://doi.org/10.1007/s10648-011-9175-6

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *The American Psychologist*, *56*(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Gruder, C. L. (1971). Determinants of social comparison choices. *Journal of Experimental Social Psychology*, *7*(5), 473–489. https://doi.org/10.1016/0022-1031(71)90010-2

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*, 210–219. https://doi.org/10.1016/j.learninstruc.2018.07.007

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168. https://doi.org/10.1016/j.lindif.2015.01.008

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*(8), 1163. https://doi.org/10.1037/a0039440

Heath, N. L., & Glen, T. (2005). Positive illusory bias and the self-protective hypothesis in children with learning disabilities. *Journal of Clinical Child & Adolescent Psychology*, *34*(2), 272–281. https://doi.org/10.1207/s15374424jccp3402_6

Honaker, J., King, G., & Blackwell, M. (2011). *AMELIA II: A program for missing data*.

Hoza, B., Pelham, W. E., Jr., Dobbs, J., Owens, J. S., & Pillow, D. R. (2002). Do boys with attention-deficit/hyperactivity disorder have positive illusory self-concepts? *Journal of Abnormal Psychology*, *111*(2), 268–278. https://doi.org/10.1037/0021-843X.111.2.268

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, *97*(1), 156–170. https://doi.org/10.1037/a0015558

Lafontaine, D., Dupont, V., Jaegers, D., & Schillings, P. (2019). Self-concept in reading: Factor structure, cross-cultural invariance and relationships with reading achievement in an international context (PIRLS 2011). *Studies in Educational Evaluation*, *60*, 78–89. https://doi.org/10.1016/j.stueduc.2018.11.005

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

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

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology*, *108*(2), 256–273. https://doi.org/10.1037/edu0000059

Marsh, H. W., & Basarkod, G. (2023). Big-fish-little-pond effect (BFLPE): Universality of psychological comparison processes. In M. Bong, S. Kim, & J. Reeve (Eds.), *Motivation science: Controversies and insights*. 100–107. Oxford University Press. https://doi.org/10.1093/oso/9780197662359.003.0010

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. *The American Psychologist*, *58*(5), 364–376. https://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., Kuyper, H., Morin, A. J., 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. https://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., & O'mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality & Social Psychology Bulletin*, *34*(4), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213

Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020). Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects: Pan-human cross-cultural generalizability over 68 countries. *European Journal of Personality*, *34*(2), 180–202. https://doi.org/10.1002/per.2232

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., & Seaton, M. (2015). The big-fish–little-pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 2, pp. 127–184). Elsevier. https://doi.org/10.1016/bs.adms.2015.05.002

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*(3), 631–669. https://doi.org/10.3102/0002831207306728

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., Van Zanden, B., Parker, P. D., Guo, J., Conigrave, J., & Seaton, M. (2019). Young women face disadvantage to enrollment in university STEM coursework regardless of prior achievement and attitudes. *American Educational Research Journal*, *56*(5), 1629–1680. https://doi.org/10.3102/0002831218824111

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

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. https://doi.org/10.1037/a0027697

OECD. (2014). *PISA 2012 technical report*. OECD.

OECD. (2019). *PISA 2018 technical report*. OECD.

Parker, P., Dicke, T., Guo, J., Basarkod, G., & Marsh, H. (2021). Ability stratification predicts the size of the big-fish-little-pond effect. *Educational Researcher*, *50*(6), 334–344. https://doi.org/10.3102/0013189X20986176

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*(1), 29–48. https://doi.org/10.1080/01443410.2013.797339

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Robitzsch, A., & Grund, S. (2019). *miceadds: Some additional multiple imputation functions, especially for 'mice' (version 3.10-28)*. https://CRAN.R-project.org/package=miceadds.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons. https://doi.org/10.1002/9780470316696

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. https://doi.org/10.1037/edu0000063

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. https://doi.org/10.1037/a0013838

Seaton, M., Marsh, H. W., Yeung, A. S., & Craven, R. (2011). The big fish down under: Examining moderators of the 'big-fish-little-pond'effect for Australia's high achievers. *Australian Journal of Education*, *55*(2), 93–114. https://doi.org/10.1177/000494411105500202

Segall, M. H., Lonner, W. J., & Berry, J. W. (1998). Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioral research. *The American Psychologist*, *53*(10), 1101. https://doi.org/10.1037/0003-066X.53.10.1101

Sewasew, D., & Koester, L. S. (2019). The developmental dynamics of students' reading self-concept and reading competence: Examining reciprocal relations and ethnic-background patterns. *Learning and Individual Differences*, *73*, 102–111. https://doi.org/10.1016/j.lindif.2019.05.010

Shankweiler, D., Lundquist, E., Dreyer, L. G., & Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing*, *8*(3), 267–294. https://doi.org/10.1007/BF00420279

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Wang, Z., & Bergin, D. A. (2017). Perceived relative standing and the big-fish-little-pond effect in 59 countries and regions: Analysis of TIMSS 2011 data. *Learning and Individual Differences*, *57*, 141–156. https://doi.org/10.1016/j.lindif.2017.04.003

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3