This MS is the final prepublication (open access) version of the published article:

Marsh, H., & Graham, Steve. (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.

Abstract

Given that the Big-Fish-Little-Pond-Effect, the negative effect of school-average achievement on academic self-concept, is one of the most robust findings in educational psychology (Marsh, Seaton et al., 2007), this research extends the theoretical model, based on social comparison theory, to study relative year in school effects (e.g., being 1 school year ahead or behind same-age students) for math constructs in PISA2003 (276,165 15-year-old students from 10,274 schools across 41 countries). The effects on academic self-concept were negative for de facto acceleration (e.g., starting early or skipping grades) and positive for de facto retention (e.g., starting late or repeating grades). These negative effects of relative year in school were: (a) cross-culturally robust across all 41 PISA2003 countries (significantly negative in most, not significantly positive in any); (b) neither substantially explained nor moderated by a diverse range of control variables (e.g., gender, school starting age, repeating grades, home language, immigrant status, SES, achievement); (c) independent of the Big-Fish-Little-Pond-Effect. The negative effects of acceleration and the positive effects of retention are consistent with a priori predictions based on frame-of-reference research, one area of social comparison research, but apparently inconsistent with some popular beliefs in relation to policy/practice based on these variables.

Cross-Cultural Generalizability of Year in School Effects: Negative Effects of Acceleration and Positive Effects of Retention on Academic Self-Concept

Self-concept is one of the oldest constructs in psychology, a major focus in many disciplines, and central to the positive psychology revolution, which focuses on how healthy, normal, and exceptional individuals can get the most from life (e.g., Diener, 2000; Marsh & Craven, 2006; Seligman & Csikszentmihalyi, 2000). Based on an integration of 17 different conceptual definitions, Shavelson, Hubner, and Stanton (1976) broadly defined self-concept as a person's self-perceptions formed through experience with, and interpretations of, one's environment. These self-perceptions are influenced especially by evaluations by significant others, by reinforcement, and by attributions for one's behavior. Self-concept is not an entity within the person, but a hypothetical construct that is important both as an outcome and as a mediating variable that helps to understand other outcomes. Self-concepts influence the way one acts and behaves, which in turn influences one's self-perceptions. In educational psychology research, academic self-concept refers to student self-perceptions of their academic self-concept in specific disciplines (e.g., math self-concept) or academic areas more generally (a global academic self-concept; Marsh, 2007). There is ample research evidence to show that a positive academic self-concept is both a highly desirable goal and a means of facilitating subsequent academic achievement, academic accomplishments, and educational choice behaviors, including subject choice, coursework selection, academic persistence, and long-term educational attainment (e.g., Chen, Yeh, Hwang, & Lin, 2013; Guay, Larose, & Boivin, 2004; Guay, Marsh & Boivin, 2003; Marsh, 1991, 2007; Pinxten, de Fraine, van Damme, & D'Haenens, 2010).

Many theoretical models (e.g., social comparison theory; Huguet, Duman et al., 2009; Marsh, Seaton et al., 2008) posit that students compare their own academic accomplishments with those of their classmates, as one basis for formation of their academic self-concept. Thus, the academic accomplishments of classmates form a frame of reference or standard of comparison that students use to form their own academic self-concepts. Based on this theoretical perspective, the Big-Fish-Little-Pond-Effect (BFLPE) posits that the effect of school-average achievement on academic self-concept is negative. More specifically (see Figure 1A), academic self-concept is positively affected by individual achievement (i.e., more able students have higher academic self-concepts); the path from individual achievement to individual academic self-concept is substantial and positive (++ in Figure 1A). However, academic self-concept is negatively affected by school-average achievement (i.e., the same student will have a lower academic self-concept when school—average achievement is high); the path from school-average achievement is negative. Although such findings were initially seen as possibly paradoxical, support for the BFLPE is now one of the most robust findings in educational psychology (e.g., Marsh, Abduljabbar et al., 2014, 2015; Marsh, Kuyper et al. 2014; Marsh, Seaton et al., 2008, in press; Nagengast & Marsh, 2012).

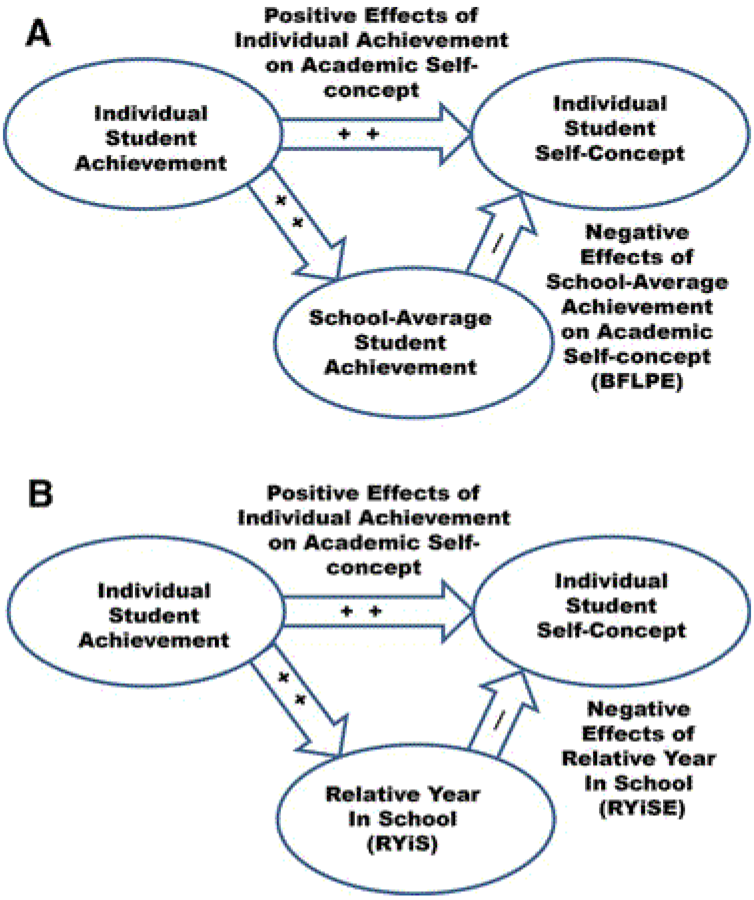

*Figure 1. (A) Conceptual model of the Big-Fish-Little-Pond-**Effect**. (Adapted with permission from Marsh, H. W. [2007]. Self-concept theory, measurement and research into practice: The role of self-concept in Educational Psychology. Leicester, United Kingdom: British Psychological Society.) (B) Conceptual Model the Relative Year in School Effect (RYiSE). See the online article for the color version of this figure.*

However, being in a school environment with highly able students, as operationalized by school-average achievement, is not the only way in which a student's frame of reference can be altered. For a variety of reasons, such as acceleration, or starting school at an early age, students can find themselves in classes with older, more academically advanced students who form a possibly more demanding frame of reference than would same-age classmates. Similarly, because of starting school at a later age or being held back to repeat a grade, students can find themselves in classes with younger, less academically advanced students than would other same-age students. For the present purposes the effect of relative year in school(RYiSE) is operationalized in terms of the difference between the year in school for a given age-cohort and the student's actual year in school. Thus, RYiS takes on positive values when a student is one or more school years ahead of same-age students, and negative values when a student is one or more school years behind other same-age students.

Although there is much research on starting age, retention and acceleration, RYiSE has not previously been considered as a frame-of-reference effect, analogous to the BFLPE (and as posited in social comparison theory). Both the BFLPE (Figure 1A) and the RYiSE (Figure 1B) posit that students' self-concepts are formed in relation to those of other students in their school. The attributes of other students in the same school constitute the frame of reference that students use. The BFLPE looks specifically at the effect of school-average achievement on academic self-concept, whereas the RYiSE looks specifically at the effect of relative year in school—the effect of being one or more years ahead, or one or more years behind, same-age classmates. Thus, the same theoretical models underlying the BFLPE (e.g., social comparison theory) also predict that RYiS effects (RYiSEs) on academic self-concept should be negative; that the effect of acceleration (a positive RYiSE) should be negative while the effect of retention (a negative RYiSE) should be positive. The logic of these predictions is similar to those based on the BFLPE: being in a context of more able students (operationalized as a higher school-average achievement) or more advanced students (operationalized as a positive RYiS, i.e., acceleration) is predicted to have a negative effect on academic self-concept. However, because relative year in school (the basis of the RYiSE) and school-average achievement (the basis of the BFLPE) are logically distinct, the two effects should be relatively independent, a prediction tested in the present investigation. Although negative RYiSEs might seem paradoxical in relation to some "accepted wisdom" particularly in relation to retention (see subsequent discussion), tests of this prediction are not only theoretically important but also have apparently important implications for the organization of schools in relation to explicit or de facto acceleration/retention. In the present investigation, we briefly review theoretical and empirical support for the BFLPE, studies of the effects of acceleration and retention, and predictions about negative RYiSEs on academic self-concept. Then, based on PISA data, we juxtapose BFLPEs and RYiSEs across 41 countries, testing the construct validity and generalizability of each.

The Big-Fish-Little-Pond-Effect (BFLPE)

As popularized recently in Gladwell's (2013)David and Goliath: Underdogs, Misfits, and the Art of Battling Giants, there is now extensive support for the generalizability of the BFLPE across diverse samples, ages, instruments, and designs (Marsh, Seaton et al., 2008; Nagengast & Marsh, 2012); this makes it one of the most robust findings in educational psychologyresearch. Indeed, based on studies of successive PISA data collections, Seaton, Marsh and Craven (2009; also see Nagengast & Marsh, 2012) have claimed support for the universality of the BFLPE as a pan-human theory. Cross-cultural research provides an important basis for testing the universality of theoretical models (Segall, Lonner, & Berry, 1998). Ideally, cross-cultural tests of generalizability would necessitate data from many countries, based on comparable samples and on measures that are equally appropriate to the different cultures. Addressing these challenges, there is strong support for the generalizability of the BFLE, based on large, nationally representative samples from PISA and from Trends in Mathematics and Science Survey data (see review by Marsh, Seaton et al., 2008; also see Nagengast & Marsh, 2012; Marsh, Abduljabbar et al., 2014; Seaton, Marsh, & Parker, 2013). Summarizing the three BFLPE-PISA studies, Nagengast and Marsh (2012) report that the effect of school-average achievement was negative in all but one of the 123 samples across the three studies, and significantly so in 114 samples. The average effect size across all 123 samples was −.223. Based on Trends in Mathematics and Science Survey data, Marsh, Abduljabbar et al. (2014) reported that the BFLPE was significantly negative for each of 26 groups (nationally representative samples of 4th and 8th grade students from 13 diverse countries; 117,321 students from 6,499 classes). Based on this cross-cultural research, the BFLPE is one of the most cross-culturally robust findings in educational research.

One approach to test the generalizability of the BFLPE is to evaluate potential moderators—particularly those of sufficient strength to eliminate the BFLPE or even to change its direction (i.e., to positive effects of school-average achievement). Moderation is of course, a double-edged sword. Significant moderators contribute to understanding the nature of the BFLPE and are potentially heuristic in terms of reducing the negative consequences. However, the failure to find substantial moderators argues for the broad generalizability and robustness of the effects. Although many potential moderators have been explored, and some have been shown to be statistically significant (e.g., Huguet, Dumas et al., 2009; Jonkmann et al., 2012; Seaton, Marsh & Craven, 2009, 2010), because of the large sample sizes typically considered, the sizes of these moderation effects are typically so small as to be of little consequence; some are not even in the direction suggested by other researchers. In particular, no studies have shown moderation sufficiently strong to reverse the negative direction of the BFLPE. A detailed summary of the large number of potential moderators that have been considered is beyond the scope of this brief overview; the reader is referred to the reviews and to further discussion of this issue (e.g., Marsh, 2007; Marsh, Seaton et al., 2008; Marsh & Seaton, in press; Seaton, Marsh, & Craven, 2009, 2010).

One of the most widely studied potential moderators of the BFLPE is individual student achievement. Indeed, the theoretical debate regarding this substantive issue of whether the BFLPE is moderated by individual student achievement (e.g., Coleman & Fults, 1985; Marsh, Kuyper et al., 2014; Marsh, Seaton et al., 2008) has important policy/practice implications for gifted education research. For example, if the negative effects of the BFLPE are limited largely to less able students, as suggested by Coleman and Fults, and more able students are actually benefited by selective schooling policies, then the results argue for a more careful screening of students, rather than problems with selective schooling per se. However, according to the theoretical model underpinning the BFLPE (Marsh, 2007) the frame of reference is largely determined by class/school-average achievement, which is necessarily the same for all students within a given school or class, so that it should be similar for the brightest and the weakest students within a given class or school. Consistent with these theoretical predictions, a growing body of empirical research (Marsh, 1984; Marsh, Kuyper et al., 2014; Marsh, Seaton et al., 2008) shows those interactions between school-average and individual student achievement are consistently small or nonsignificant, and not even consistent in direction—that bright, average, and less bright students experience negative BFLPEs to a similar extent.

Year in School Effects (YiSEs)

School starting age, acceleration (i.e., skipping a year in school), and being retained (i.e., repeating a year in school) have been studied extensively in relation to academic achievement (e.g., Alexander, Entwisle, & Dauber, 2003; Jimerson, 2001; but see Reynolds, 1992; Roderick, 1994; Roderick & Engel, 2001). As noted by Jimerson and Brown (2013, p. 140), “because of potential short- and long-term effects that grade retention can have on student achievement and socioemotional outcomes, it remains a controversial topic in research and practice.” Nevertheless, there is a general belief, supported by some research evidence that, in terms of academic achievement, retention has negative effects, while acceleration may have positive effects (e.g., Hattie, 2012). However, critical design and methodological issues, such as the need for appropriate control groups and prior measures, dictate caution in reaching overarching conclusions such as these (Jimerson & Brown, 2013). Thus, for example, although accelerated students score higher on achievement tests than do same-age students who are not accelerated, they do not achieve better compared with older (same-class) students in their accelerated class; also, results for affective outcomes are inclusive (e.g., Kretschmann, Vock, & Lüdtke, 2014; Kulik & Kulik, 1984, 2004). Making a similar point based on their meta-analysis of grade retention studies that controlled quality study, Allen, Chen, Willson, and Hughes (2009)reported that their results “challenge the widely held belief that retention has a negative effect on achievement” (p. 480). Studies showing negative effects of retention are largely limited to poor quality studies that do not control sufficiently for pre-existing differences.

Consistently with the Allen et al. (2009) meta-analysis, a number of recent studies challenge the view that retention has negative effects, or else show that negative effects in prior studies are likely the result of inadequate control for selection effects (Im et al., 2013; Moser, West, & Hughes, 2012; Wu, West, & Hughes, 2008, 2010). Thus, for example, in a 4-yearlongitudinal study using propensity matching to match primary students who had been retained with unretained, promoted students, Wu et al. (2010) found that retention had short-term positive effects on school-belonging, teacher-rated engagement, and academic self-concept. In a follow-up of this same study, Im et al., found that retained and promoted students following transition to middle school did not differ in terms of achievement, engagement, or school belonging (although they did not report the follow-up measures of academic self-concept considered in the earlier study; the focus of the present investigation). Although Im et al. had expected to find negative effects of retention based on previous research (e.g., Alexander et al., 2003), they argued that their propensity matching results were stronger than the statistical adjustment procedures typically used.

In respect of the purposes of the present investigation, there has been little large-scale, rigorous research into the effects of retention and acceleration on academic self-concept and related self-belief and motivation variables, and very little research evaluating these issues from a rigorous cross-cultural perspective. Of particular relevance to the present investigation, Marsh, Chessor, Craven, and Roche (1995, Study 2) reported that the youngest participants in a gifted-and-talented intervention were accelerated nearly a year ahead of their matched comparison group classmates (matched on gender, yearin school, and prior achievement, but not on age) before the start of the intervention. They found that students in the intervention group had substantially lower self-concepts than matched (but typically older) comparison students at T1, before the start of the gifted-and-talented program and that this effect was evident for both academic and nonacademic components of self-concept. However, the decline in academic self-concepts associated with participation in the gifted-and-talented intervention did not interact with age. Thus, the apparently negative effects of acceleration on self-concept (negative RYiSEs) were apparently independent of and in addition to the negative effects of class-average achievement (the BFLPE). Nevertheless, this pattern of results may be consistent with the frame-of-reference effects that are the basis of the BFLPE, in that being placed in a context with older students who are physically, socially, emotionally, and academically more mature than would be the case if the same student were in classes with same-aged classmates, may alter the frame of reference that accelerated students use to form their self-concepts in a manner that is consistent with the theoretical basis of the BFLPE. Although their research was not designed specifically to evaluate this question, Marsh et al. (1995) suggested that these results could offer a potentially important direction for further research. Of interest to those authors, although not a focus of the present investigation, Marsh et al. (1995) found that this effect might not be domain-specific: it may generalize to nonacademic as well as academic domains of self-concept.

The Present Investigation

In most school systems throughout the world, students typically are grouped into the same grade or year in schoolaccording to their age, rather than to their abilities in general, or in particular school subjects. Thus, with the exception of students who start school early or late, those identified as gifted or in need of remedial assistance, it is typical for students within the same class to be of a similar age. For example, in the present investigation, based on nationally representative samples of 15-year-olds (total N = 276,165) from 41 countries

(PISA2003 data, OECD, 2005a), 67% of the students were in their modal year in school for their country. However, for nearly all countries, there were 15-year-old students accelerated one or more years relative to the modal year in school (e.g., students in Years 11 or 12 when their modal or “age-appropriate” year group was Year 9 or 10), whereas others were in year groups one or more years less than their modal year group (e.g., students in Years 7 or 8 when their modal or “age-appropriate” year group was Year 9 or 10). Indeed, sampling students of a particular age, rather than according to year in school, is a defining feature of the PISA data. Typically, this issue is considered a potentially complicating feature of PISA data that needs to be controlled for when considering other issues (e.g., including year in school as a covariate; Nagengast & Marsh, 2013). However, this feature has seldom been a primary emphasis of research, particularly in relation to psychosocial variables, the focus of the present investigation. This design feature of PISA does however provides an ideal opportunity to test the Marsh et al. (1995)speculations based on apparently independent, negative frame-of-reference effects based on school-average achievement and RYiS. Although the research hypotheses in relation to the BFLPE have extensive support from previous research, those in relation to RYiSEs are apparently new, as they have not previously been proposed or tested.

Because of the complex nature of the present investigation, this outline of the present investigation begins with a brief overview of the conceptual research design and of the variables considered (Figure 2; also see Supplemental Materials, Section 1, for a more detailed description of the variables to be considered), followed by a statement of the specific research hypotheses and research questions that are pursued. The central results are the effects of the main independent variables (particularly school-average achievement and relative year in school, RYiS) on the main dependent variable (math self-concept, M-ASC). The main findings in relation to the negative effects of school-average achievement follow closely from previous research already discussed (also see Figure 1A).

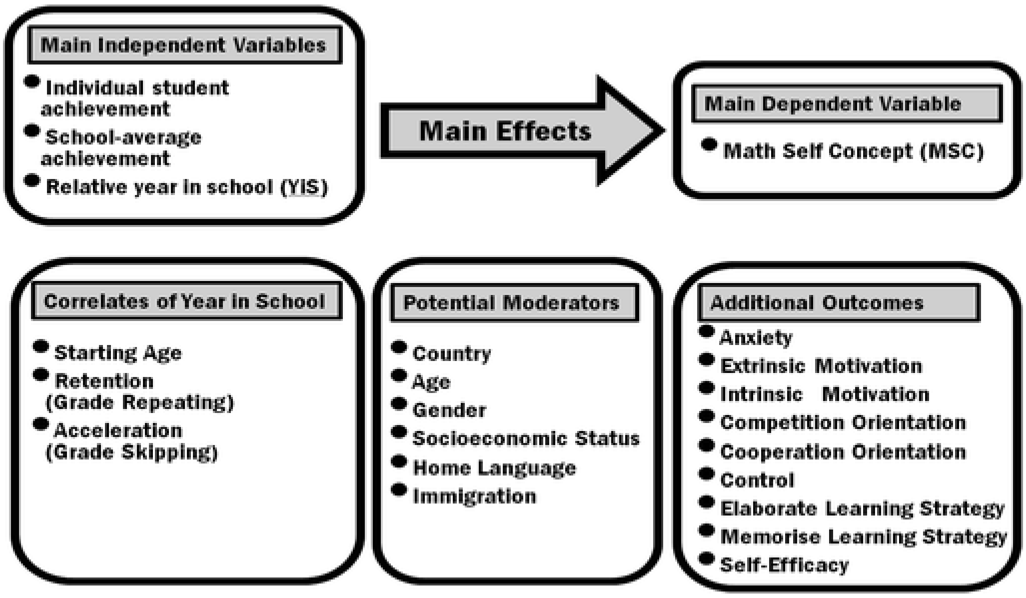

Figure 2. Conceptual design of the present investigation. The critical results are the main effects of the main independent variables (particularly school-average achievement and year in school) on the main dependent variable (math self-concept). Correlates of year in school are used to test the assumption that year in school captures all or much of the variance that can be explained by these correlates. Potential moderators are used to test if

any of these variables substantially moderate the effects of school-average achievement (the BFLPE) or year in school on math self-concept. Additional outcomes are used to test the effects of school-average achievement and year in school on these outcomes (and how much of these effects can be explained in terms of math self-concept); analyses based on these variables are presented in Supplemental Materials, Section 6, but only summarized briefly in the printed version of the article. Definitions of the variables and the wording of the items are presented in detail in the Supplemental Materials, Section 1.

However, this theoretical perspective on RYiSEs has apparently never been considered previously. This is also the case for the generalizability of the results over 41 different countries, for potential moderators of RYiSEs, and perhaps even the predicted direction of the effects on M-ASC. Variables in the box labeled "correlates of year in school" are used to test the assumption that year in school captures all or much of the variance that can be explained by retention, acceleration, and starting age. Variables in the box labeled "potential moderators" are used to test with if any of these variables moderate the effects of school-average achievement (the BFLPE) or RYiSEs on M-ASC. Variables in the box labeled "additional outcomes" are used to test the effects of school-average achievement and year in school on these outcomes (although actual analyses are included in Supplemental Materials, Section 6). Definitions of the variables and the wording of the items are given in detail in the Supplemental Materials, Section 1. Based on this research design and set of variables, two overarching hypotheses were posited that guided the statistical analyses and presentation of results.

Research Hypotheses

Hypothesis 1: A basic frame-of-reference model. After controlling for individual achievement there will be negative effects on M-ASC associated with school-average achievement (the negative BFLPE) and year-in-school (i.e., negative RYiSEs) that are reasonably independent of each other (see Figure 1A). The juxtaposition of Figures 1 and 2 demonstrates that this basic model is essentially the same as that which has been used extensively in BFLPE research, but with the critical addition of RYiS.

Hypothesis 2: Cross-cultural generalizability. Both these negative effects on M-ASC (the BFLPE and the negative RYiSEs) will be reasonably consistent across OECD and non-OECD countries, and across the entire set of 41 countries more generally.

Research Questions

In addition to these research hypotheses two research questions were pursued, where there was no clear basis for making a priori hypotheses (based on prior theory or prior empirical research), but that nevertheless indicate interesting issues to pursue, with theoretical, substantive or policy-related implications.

Research question 1

How are the BFLPEs and, particularly, the negative RYiSEs, related to whether students begin school early or late, repeat a year, or are accelerated? It was anticipated that all these variables would be relatively unrelated to the BFLPE but may further explicate processes underlying the hypothesized negative RYiSEs. In particular, based on the assumption that the effects of retention, acceleration, and starting age can all be largely encapsulated into effects of RYiS, it was anticipated that these correlates of RYiS (see Figure 2) will have little or no systematic effect beyond what can be explained by RYiS. This finding would have very important implications, providing an important bridge between different areas of research where there has been surprisingly little cross-fertilization, and providing a powerful tool for understanding the effects of retention, acceleration, and school-starting age.

Research question 2

How are the BFLPE and RYiSEs related to or moderated by background characteristics: achievement, gender, age, immigrant status, home language, and various components of socioeconomic status? Consistently with previous research it is anticipated that interactions between these variables and school-average achievement (the BFLPE) will all be small, but leave as a research question whether any interactions with RYiS are meaningfully large (e.g., sufficiently large so that negative RYISEs are found to be positive for some students).

Method

Sample and Variables

The data for the present investigation are from the Program of International Student Assessment, which is administered to 15-year-old students from countries all over the world every 3 years. In the PISA2003 data used here, the primary focus was on math. Students completed paper-and-pencil tests to assess their knowledge and skills in reading, math, science, and problem solving. In addition, each student completed a questionnaire that assessed student and family background variables, and a variety of psychosocial variables, including math self-concept. In the present investigation, the sample consisted of 276,165 15-year-old students, in 10,274 schools, across 41 countries. Support for the reliability and validity of the achievement scores is presented in the PISA2003 technical manuals (Organisation for Economic Cooperation and Development, 2005a, 2005b).

The primary outcome variable in the present investigation is M-ASC, measured with five items (I am just not good at mathematics; I get good marks in mathematics; I learn mathematics quickly; I have always believed that mathematics is one of my best subjects; and In my mathematics class, I understand even the most difficult work), was based on the IRT scales index, with a median reliability of .89. This IRT score was provided with the PISA data and the use of this score is recommended in the PISA manual (Organisation for Economic Cooperation and Development, 2005b). All questionnaire scale scores, including M-ASC, were scaled by PISA to have Mn = 0 and *SD* = 1 across Organisation for Economic Cooperation and Development (OECD) countries (see Schulz, 2005, for further description of scaling).

RYiS was defined as the deviation between each student's actual year in school and the mean year in school for the student's country. As noted in the PISA manual, across the entire sample, 67% of the students were in the modal year; 20.7%, 3.7% and 1.1% were 1, 2, or 3 years, respectively, behind the modal value, and 7.1%, 0.4% and 0.01% were 1, 2, or 3 years, respectively, ahead of the modal year.

Math achievement in PISA is intended to represent the extent to which 15-year-olds have acquired some of the knowledge and skills that are essential for full participation in society and to make practical use of mathematics in different situations relative to adult life, rather than what is taught in particular countries. The assessment consists of paper-and-pencil tests using a variety of item types designed to measure a broad range of mathematical skills, competence levels, and types of application. PISA math achievement scores are normed to have a M = 500, SD = 100 across all OECD countries, with a test reliability = .845; Organisation for Economic Cooperation and Development, 2005b; for further discussion, see Organisation for Economic Cooperation and Development, 2005a, 2005b). For the present purposes, math achievement was defined by the first plausible value that is available as part of the public PISA database and is an unbiased "representation of the range of abilities a student might reasonably have . . . Instead of directly estimating a student's ability" (Organisation for Economic Cooperation and Development, 2005a, p. 75).

I began by standardizing M-ASC, math achievement and all other individual student scores (*M* = 0, *SD* = 1) across the entire sample, to facilitate interpretations. Following previous research showing quadratic relations between M-ASC and math achievement (e.g., Nagengast & Marsh, 2012, 2013) I then created the quadratic components for math achievement by squaring the standardized math achievement score for each student. School-average achievement was computed as the mean of the standardized achievement score within each school. Finally, cross-product terms were created to test moderation effects for selected variables, multiplying each by RYiS and school-average achievement. None of these constructed cross-product variables (quadratic student achievement, school-average achievement, and cross-product terms involving relative year in school and school-average achievement) were restandardized; thus, keeping them in the same metric as individual student variables. Additional variables (see Supplemental Materials, Section 1, for more detail) were then considered, to evaluate other research hypotheses and questions about the nature and construct validity of the BFLPEs and RYiSEs. These included: student's age, age when first started school, number of repeated grades, different components of family socioeconomic status, immigrant status, and home language.

Analysis

In the last quarter century, multilevel modeling has become a central research tool in the social sciences, and has had a profound effect on BFLPE research. Ignoring a hierarchical structure can result in serious statistical problems, not only violating assumptions of independence but also increasing the likelihood of finding statistical significance where none exists (Hox, 2002). Thus, because most educational data has a multilevel structure (e.g., students nested within classes and schools), multilevel analyses are particularly important in educational research, even when researchers are only interested in relations among variables at the individual student level. However, a major advantage of multilevel modeling over single level analysis is in the ability to explore appropriately relationships among variables at different levels (Goldstein, 2003; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). In the typical application of multilevel modeling, outcome variables are related to several predictor variables at the individual level (e.g., students) and at the group level (e.g., classes, schools). In this literature, models that include the same variable at both the individual level and the aggregated group level are called contextual analysis models (Marsh, Lüdtke et al., 2013; Raudenbush & Bryk, 2002). The central question in such contextual studies is whether the aggregated group characteristic has an effect on the outcome variables, after controlling for variables at the individual level. In contextual studies the critical question is the relative sizes of the effects of individual and group-average constructs in predicting relevant outcome measures when both individual and group-average variables are included in the analysis. In this respect the BFLPE paradigm is a classic contextual study in which individual and school-average achievement are used to predict academic self-concept and the appropriate statistical analysis involves multilevel modeling (Marsh, Lüdtke et al., 2009). The PISA data have a three-level hierarchical structure: students (Level 1) nested within schools (Level 2), and schoolsnested within countries (Level 3). Consequently, to accommodate this hierarchical structure, a multilevel modeling program (MlwiN) was used to analyze these data (see Rasbash, Steele, Browne, & Prosser, 2004 for details). A multilevel regression equation consists of two parts: a fixed component and a random component. Consistent with earlier studies (Marsh & Hau, 2003; Seaton, Marsh, & Craven, 2009, 2010; also see Marsh, Seaton et al., 2008), the fixed component included effects of individual achievement (both linear and quadratic) and school-average achievement, but also RYiS, the focus of the present investigation, and an apparently new contribution to the study of frame-of-reference effects (see Figure 2). The random component consists of the intercepts, variances, and covariances of the levels used in the model. In this case, the intercepts of country (Level 3), school (Level 2), and student levels (Level 1) were of particular interest, as they showed how much the intercepts of the regression equations varied between countries, between schools, and between students. Additionally, multilevel modeling can take into account that regression equations of nested data can vary, not only in their intercepts, but also in their slopes. Thus, I allowed selected variables—particularly RYiS and school-average achievement—to vary at the country level, to evaluate the extent of country-to-country variation in their effect.

Results

Basic Models: The Negative Effects of School-Average Achievement and Year in School (Hypothesis 1)

I begin by evaluating initial support for Hypothesis 1: that the effects of school-average achievement (BFLPE) and relative year in school (RYiSE) on M-ASC would both be negative. These are evaluated with three models (Models 1A–1C in Table 1), which consider each of these predictor variables separately and in combination. Nevertheless, the largest effect in each of these models, not surprisingly, is student achievement: higher achieving students have higher M-ASCs. However, although there is a substantial positive linear effect of individual student achievement (varying from .510 to .557 in different models), there is also a small quadratic component that is highly significant from a statistical perspective (see the first panel in Figure 3).

Table 1

Prediction of Math Self-Concept: Basic Model of the Negative Effect of School-Average Achievement (BFLPE) and Relative Year in School (Models 1A–C) and Their Generalizability Over 41 Countries (2A) and Organization for Economic Cooperation and Development (OCED) vs. Non-OECD Countries (2B)

Effect	Model 1A β (*SE*)	Model 1B β (*SE*)	Model 1C β (*SE*)	Model 2A[a] β (*SE*)	Model 2B β (*SE*)
Fixed effects					
Main effects					
Student Linear ACH (SLA)	.510 (.002)	.540 (.003)	.557 (.003)	.549 (.019)	.549 (.019)
Student Quad ACH (SQA)	.103 (.002)	.104 (.006)	.099 (.002)	.107 (.006)	.107 (.006)
Year in School (YiS)	−.104 (.003)		−.081 (.002)	−.095 (.006)	−.097 (.009)
School ACH (BFLPE)		−.370 (.006)	−.342 (.006)	−.304 (.024)	−.311 (.023)
OECD					−.039 (.033)
Interaction effects					
OECD × RYiS					−.009 (.008)
OECD × BFLPE					−.037 (.021)
Random effects					
Level 3 intercept	.097 (.022)	.037 (.008)	.044 (.010)	.059 (.016)	.059 (.016)
Student Linear ACH				.015 (.003)	.015 (.003)
Student Quad ACH				.001 (.000)	.001 (.000)
Year in School (YiS)				.003 (.001)	.003 (.001)
School ACH (BFLPE)				.021 (.007)	.021 (.007)
Level 2 intercept	.057 (.001)	.034 (.001)	.035 (.001)	.029 (.002)	.029 (.002)
Level 1 intercept	.783 (.002)	.785 (.002)	.782 (.002)	.774 (.002)	.774 (.002)

Note. ACH = achievement; RYiS = relative year in school, the difference between starting age and mean starting age of the country; critical parameters are the effects of grade and school-average achievement (BFLPE), shaded in grey. All parameter estimates (βs) are significant at .05 level when they differ from zero by more than 2 *SE*s.

[a] In supplemental analyses, I fitted a model in which all 10 residual covariances at the country level were estimated. However, inclusion of these additional parameters had almost no effect on other parameter estimates, and all 10 residual covariances were very small; despite the very large *N*, 7 of 10 were nonsignificant ($p < .05$; none at $p < .01$). Nevertheless, the residual covariance between the BFLPE and RYiSEs (.0030, $SE = .0015$) was marginally significant.

Prediction of Math Self-Concept: Basic Model of the Negative Effect of School-Average Achievement (BFLPE) and Relative Year in School (Models 1A–C) and Their Generalizability Over 41 Countries (2A) and Organization for Economic Cooperation and Development (OCED) vs. Non-OECD Countries (2B)

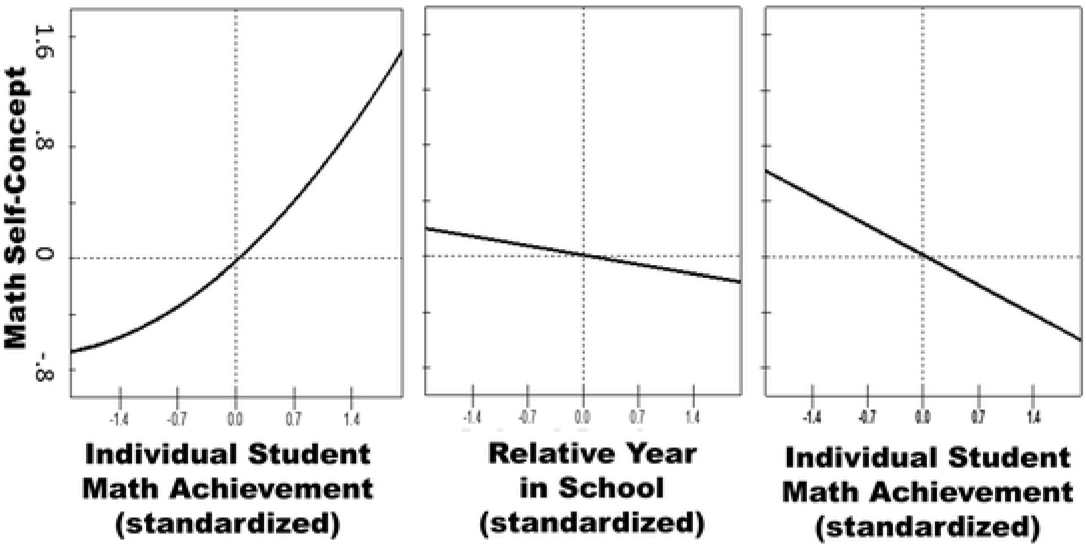

Figure 3. The relation of Math Self-Concept to Individual Achievement, the negative effect of year in school (standardized school grade), and the negative effect of school-average achievement (BFLPE, negative Big-Fish-Little-Pond-Effect). Note. Figure based on predicted math self-concept values for predictor values in the range of −2 to 2 SDs from the mean (0) of each of the predictor variables.

In Model 1A, the negative RYiSE (−.104) is significantly negative when the BFLPE is not considered: 15-year-old students have lower M-ASCs when they are in higher RYiS levels. In Model 1B, the negative effect of school-average achievement, the BFLPE (−.370), is highly significant when RYiS is not considered: students in schools where the average achievement level is higher have lower M-ASCs. In Model 1C (see Table 1), which contains both school-average achievement and RYiS, the effects of each of these variables are similar in size and direction to Models 1A and 1B, but slightly smaller. In Model 1C the negative effect of school-average achievement (BFLPE = −.342, *SE* = .006) is more negative than the negative effect of relative year in school (RYiSE = −.081, *SE* = .002). While each of the effects is highly significant from a statistical perspective, the difference between the two effects is in itself statistically significant. Consistent with these results, the residual variance terms for school-average achievement and RYiS were positively related but very small, and only marginally significant (residual covariance = .0030, *SE* = .0015; see footnote in Table 1). Thus, the negative effects of school-average achievement (the BFLPE) and relative year in school (RYiSEs) are nearly independent. In Figure 3, 1 juxtapose the positive effects of individual student achievement with the negative effects of RYiS and the BFLPE. These demonstrate, as shown in Table 1, that the negative effect of the BFLPE is more negative (i.e., the negative slope is steeper) than the negative RYiSE.

Cross-Cultural Generalizability of the Negative Effects of School-Average Achievement and RYiS Level (Hypothesis 2)

Having demonstrated the negative effects of both school-average achievement and RYiS (Hypothesis 1), I now move to the issue of the cross-cultural generalizability of these effects across OECD and non-OECD countries, and across the 41 countries more generally (Hypothesis 2).

In Model 2B (see Table 1), I first evaluate whether the negative RYiSEs or the negative BFLPEs on M-ASC differ for OECD and non-OECD countries. Of critical interest are the interactions of the dichotomous OECD variable with the RYiSE and BFLPE. However, as both these are nonsignificant, the results indicate that neither the negative RYiSEs nor the negative BFLPE differ significantly for OECD and non-OECD countries. In Model 2A (see Table 1), the effects on M-ASC of individual student achievement, relative year in school (RYiSE), and school-average achievement (the BFLPE) are all made random at the country level; thus, allowing the effects to vary from country to country. Due in part to the large numbers, country-to-country variation is statistically significant for each of these effects. Nevertheless, the residual variance components are all small, indicating considerable consistency in the effects across the 41 countries. However, of particular interest in relation to Hypothesis 2 are the residual variance components (see Table 1) for relative year in school (RYiSEs) and school-average achievement (the BFLPE) that I now explore in more detail.

Cross-cultural generalizability of the negative effects of RYiS level

In Model 2A, the residual variance component for RYiS (.003) is small, relative to the negative RYiSE (−.095). Also, as shown in the corresponding "caterpillar" plot (see Figure 4), as well as in the separate listings of effects for each of the 41 countries (see Table 2), the confidence intervals around the RYiSEs tend to be small. For example, in the caterpillar plot for the RYiSE, each of the 41 countries is ranked in terms of the size of the RYiSE, and a confidence interval (±1 *SE*) around the estimated effect of each country. This plot (also see Table 2) shows that while the effect is consistently negative, for several countries the estimated effect is within one standard error of zero. However, supplemental analyses showed that for countries where the RYiSE was small, there was also little or no variation in RYiS level (e.g., for Iceland, Japan, and Yugoslovia, nearly all 15-year-olds were in the same year in school), so that it is not surprising that the RYiSE was small. Nevertheless, the caterpillar plots shows that at least the direction of the RYiSE is consistent across the set of countries, in that it was not significantly positive in any of the 41 countries.

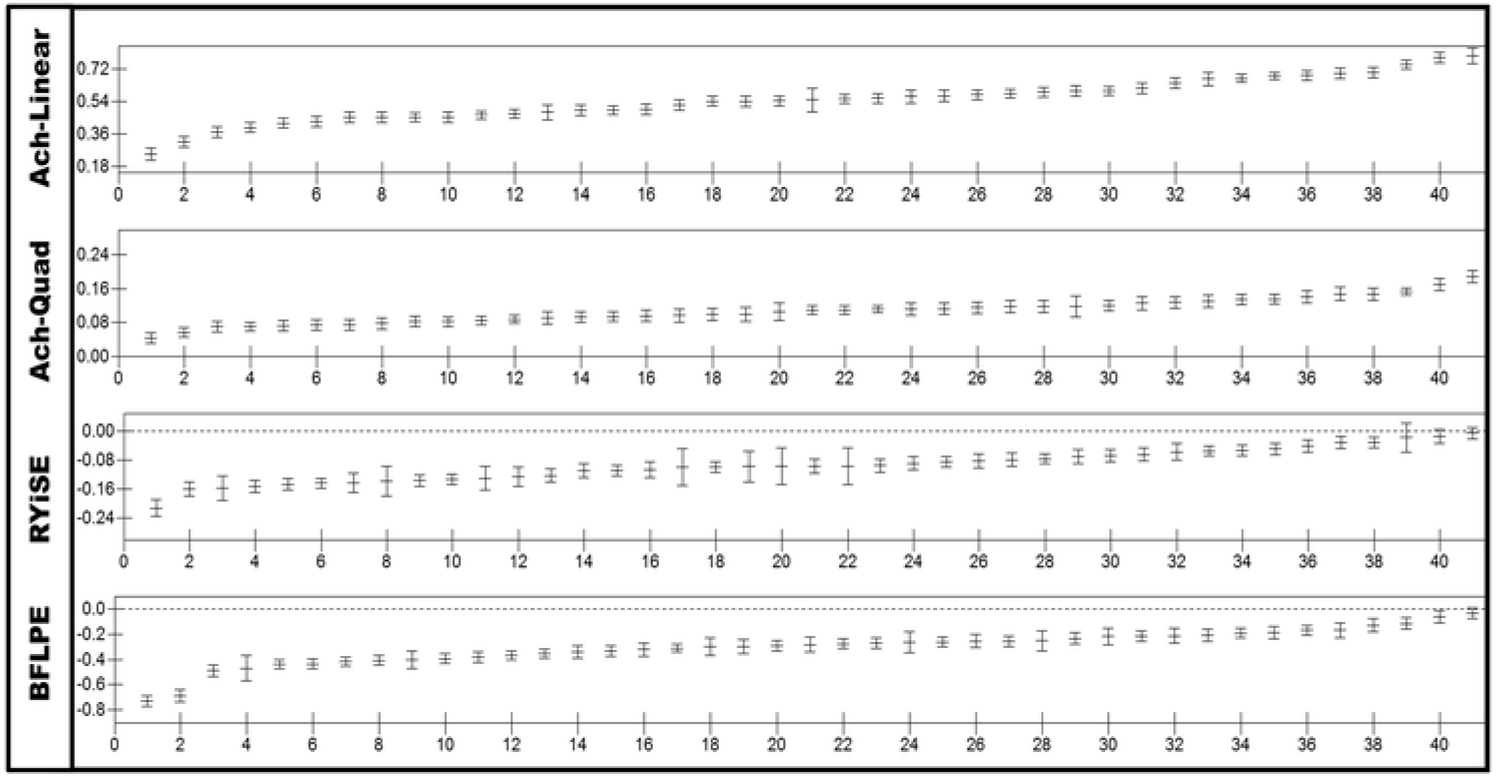

Figure 4. *Caterpillar plots showing the country-to-country variation in the effects on math self-concept of individual student achievement (linear and quadratic), RYiSE (relative year in school effect), and the BFLPE (Big-Fish-Little-Pond-Effect, negative effect of school-average achievement) for each of 41 countries. For each plot the set of 41 countries is rank-ordered in terms of the predicted value (i.e., the lowest ranked countries are not necessarily the same for each plot). For each country the predicted mean effect and confidence interval (±1 SD) are shown. For individual student achievement, all predicted values are positive. For RRYiSEs and BFLPEs, all predicted values are negative Ach = achievement.*

Table 2
Parameter Estimates and SEs For 41 Countries

Country	L1Ach	*SE*	$L1Ach^2$	*SE*	RYiS	*SE*	BFLPE	*SE*
Australia	.467	.022	.087	.009	−.133	.014	−.277	.038
Austria	.615	.030	.115	.014	−.071	.019	−.490	.043
Belgium	.455	.025	.084	.010	−.137	.016	−.441	.034
Brazil	.568	.036	.082	.013	−.006	.015	−.380	.042
Canada	.682	.021	.112	.008	−.144	.013	−.369	.032
Czech Republic	.601	.028	.074	.012	−.097	.019	−.435	.038
Denmark	.686	.027	.090	.014	−.142	.027	−.297	.056
Finland	.697	.029	.147	.014	−.213	.023	−.299	.071
France	.572	.031	.118	.015	−.057	.024	−.342	.049
Germany	.784	.030	.095	.013	−.153	.017	−.731	.041
Greece	.542	.027	.077	.013	−.015	.019	−.117	.047
Hong Kong	.454	.030	.094	.012	−.109	.014	−.255	.042
Hungary	.455	.029	.099	.014	−.065	.018	−.296	.042
Iceland	.701	.028	.147	.016	−.096	.051	−.253	.082
Indonesia	.249	.034	.070	.012	−.077	.014	−.195	.038
Ireland	.455	.028	.125	.016	−.031	.015	−.170	.061
Italy	.601	.024	.070	.010	−.041	.016	−.418	.033
Japan	.371	.028	.043	.013	−.101	.051	−.334	.041
Korea	.428	.029	.092	.013	−.098	.043	−.034	.043
Latvia	.522	.027	.112	.014	−.109	.020	−.255	.054
Liechtenstein	.549	.064	.118	.025	−.138	.041	−.472	.101
Luxembourg	.499	.028	.142	.015	−.124	.018	−.403	.071
Macao (China)	.482	.042	.105	.021	−.107	.022	−.266	.085
Mexico	.670	.023	.153	.009	−.055	.013	−.310	.031
Netherlands	.667	.037	.100	.017	−.161	.020	−.688	.045
New Zealand	.420	.027	.120	.013	−.126	.026	−.323	.051
Norway	.745	.026	.188	.014	−.098	.051	−.216	.066
Poland	.559	.026	.170	.014	−.131	.033	−.284	.059
Portugal	.643	.028	.130	.015	−.080	.018	−.132	.051
Russia	.397	.025	.072	.012	−.096	.017	−.215	.043
Slovakia	.585	.025	.074	.012	−.082	.019	−.394	.039
Spain	.555	.023	.110	.011	−.049	.016	−.262	.041
Sweden	.578	.025	.135	.013	−.158	.035	−.214	.059
Switzerland	.495	.024	.109	.011	−.147	.017	−.407	.036
Thailand	.318	.030	.112	.014	−.089	.018	−.189	.046
Tunisia	.794	.042	.118	.015	−.084	.015	−.062	.051
Turkey	.542	.029	.056	.012	−.032	.017	−.272	.040
United Kingdom	.474	.024	.135	.011	−.100	.015	−.356	.037
United States	.544	.025	.127	.013	−.069	.017	−.235	.044
Uruguay	.593	.028	.082	.011	−.054	.015	−.171	.040
Yugoslavia	.492	.030	.097	.015	−.018	.040	−.208	.049
Total Estimate	.549	.033	.107	.006	−.095	.006	−.304	.024
Total Resid Var	.015	.016	.001	.000	.003	.001	.021	.007

Note. L1Ach = linear effect of student level achievement; $L1Ach^2$ = quadratic effect of student level achievement; RYiS = relative RYiS; BFLPE = Big-Fish-Little-Pond-Effect (negative effect of school-average achievement). All parameter estimates are significant at .05 level when they differ from zero by more than 2 *SE*s.

Parameter Estimates and SEs For 41 Countries

Cross-cultural generalizability of the negative BFLPE

In Model 2A, the residual variance component for the BFLPE (.029) is small, relative to the negative BFLPE (−.304). Again, turning to the corresponding caterpillar plot (Figure 4; also see Table 2), the estimated BFLPEs for all 41 countries are negative; in only one country (Korea) was the negative effect less than 1 *SE* below the mean. However, supplemental analyses again showed that for countries where the BFLPE was relatively small, there was less school-to-school variation in the level of school-average achievement. Nevertheless, the caterpillar plots (Figure 4; also see Table 2) show that at least the direction of the

BFLPE effect is consistent across the set of 41 countries and that the BFLPE is not positive in any of the countries.

Starting Age, Repeating RYiS, and Acceleration (Research Question 1)

As emphasized earlier, students can find themselves in classes with either older or younger students, for any of a variety of different reasons—including starting age (i.e., starting school early or late), repeating one or more years in school, or acceleration (i.e., skipping one or more years in school). Here I evaluate how the negative effect of school-average achievement (the BFLPE) and, in particular, the negative effect of relative year in school (RYiSE) is related to these possibilities. In Table 3 a pair of models (labeled A and B) are tested for each covariate (e.g., starting age): initially one without RYiSE (A) and then one with RYiSE. Of particular interest are the effects of the covariates with and without the inclusion of relative year in school; how the inclusion of relative year in school changes the effects of these covariates and how the inclusion of these covariates changes the RYiSEs.

Table 3

Prediction of Math Self-Concept: Models of the Negative Effect of School-Average Achievement (BFLPE; Models 3a, 4a, 5a, and 6a) and Relative Year in School (RYIS; Models 3b, 4b, 5b, and 6b) With Inclusion of Covariates Related to Relative Year in School (Starting Age, Repeating a Year in School, Acceleration)

	Starting Age (cov)		Repeat (cov)		Acceleration (cov)		Repeat/Acceleration[a] (covs)	
Effects	Model 3a β (*SE*)	Model 3b β (*SE*)	Model 4a β (*SE*)	Model 4b β (*SE*)	Model 5a β (*SE*)	Model 5b β (*SE*)	Model 6a β (*SE*)	Model 6b β (*SE*)
Fixed effects								
Student Linear ACH (SLA)	.529 (.019)	.551 (.019)	.533 (.019)	.549 (.019)	.527 (.019)	.551 (.019)	.535 (.019)	.551 (.019)
Student Quad ACH (SQA)	.109 (.005)	.106 (.006)	.109 (.005)	.107 (.006)	.110 (.019)	.106 (.006)	.107 (.005)	.106 (.006)
Relative Year in School (RYiS)		−.096 (.009)		−.104 (.009)		−.096 (.009)		−.105 (.009)
School ACH (BFLPE)	−.322 (.023)	−.303 (.024)	−.318 (.023)	−.308 (.024)	−.323 (.024)	−.305 (.024)	−.317 (.024)	−.305 (.024)
Covariate	.020 (.003)	.001 (.003)	.026 (.002)	−.017 (.003)	−.003 (.002)	.005 (.002)	.027 (.002)	−.017 (.003)
Covariate							.000 (.001)	.005 (.002)
Random effects								
Level 3 intercept	.051 (.011)	.058 (.013)	.051 (.011)	.058 (.013)	.051 (.011)	.059 (.013)	.053 (.012)	.059 (.013)
Student Linear ACH	.014 (.003)	.015 (.003)	.014 (.003)	.014 (.003)	.014 (.003)	.015 (.003)	.014 (.003)	.015 (.003)
Student Quad ACH	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)
Relative Year in School (RYiS)		.002 (.001)		.003 (.001)		.003 (.001)		.002 (.001)
School ACH (BFLPE)	.020 (.005)	.020 (.005)	.020 (.005)	.022 (.005)	.020 (.005)	.022 (.005)	.021 (.005)	.020 (.005)
Level 2 intercept	.029 (.001)	.029 (.001)	.029 (.001)	.029 (.001)	.029 (.001)	.029 (.001)	.029 (.001)	.029 (.001)
Level 1 intercept	.778 (.002)	.774 (.002)	.779 (.002)	.775 (.002)	.780 (.002)	.775 (.002)	.780 (.002)	.776 (.002)

Note. BFLPE = Big-Fish-Little-Pond-Effect; ACH = achievement; RYiS = relative year in school, the difference between starting age and mean starting age of the country; covariates are: Age = current age; Starting Age = Age first started primary school (Year 1); Repeating = Number of repeated RYiSs; Acceleration = number of skipped RYiSs. For each covariate two models are presented, one without the RYiS variable (A) and one with the RYiS variable (B), to determine whether the effect of the covariate changes when RYiS is included. Critical parameters are the effects RYiS, school-average achievement (BFLPE), and the covariates, shaded in grey. For purposes of comparison, results with no covariates are presented in Model 1C from Table 1). All parameter estimates (βs) are significant at $p < .05$ when they differ from zero by more than 2 *SE*s.

[a] For Models 6a and 6b, two covariates were included. The first is repeating (used in Models 6a and 6b), the second is acceleration. When used together, they can be thought of as a step regression (with neither repeat nor acceleration taking on the implicit value of zero).

Prediction of Math Self-Concept: Models of the Negative Effect of School-Average Achievement (BFLPE; Models 3a, 4a, 5a, and 6a) and Relative Year in School (RYIS; Models 3b, 4b, 5b, and 6b) With Inclusion of Covariates Related to Relative Year in School (Starting Age, Repeating a Year in School, Acceleration)

Starting age

Starting age has a small but significantly positive effect on M-ASC (.020, $p < .001$; Model 4a, Table 3) when RYiS level is not considered—students who are older when they start school have slightly higher M-ASCs when aged 15. However, when relative year in school is included, the effect of starting age becomes nonsignificant. Thus, the positive effect of starting age on M-ASC can be explained in terms of relative year in school. Not surprisingly, the inclusion of starting age had almost no influence on any other predictor variables (see comparison of Models 1C and 2A in Table 1 with Model 3b in Table 3).

Repeating a year in school

Repeating a year in school has a small but significantly positive effect on M-ASC (.026, $p < .001$; Model 4a) when RYiS level is not considered: students who have previously repeated a year in school have slightly higher M-ASCs when aged 15. However, when relative year in school is included, the effect of repeating a year in school becomes slightly negative (−.017, $p < .001$; Model 4b). Thus, the positive effect of repeating a grade on M-ASC can be explained in terms of RYiS. Indeed, the inclusion of repeating a year in school resulted in a slightly more negative RYiSE (−.095 vs. −.106). Again, the inclusion of repeating a RYiS had almost no influence on any other predictor variables (see Models 1C and 2A in Table 1 vs. Model 4b).

Acceleration

Acceleration has no statistically significant effect M-ASC (−.003; Model 5a) when relative year in school is not considered. When relative year in school is included, the effect of acceleration becomes slightly positive (.002, $p < .05$; Model 5b). However, the inclusion of acceleration had almost no influence on any other predictor variables (see Models 1C and 2A in Table 1 vs. Model 5b).

In Models 6a and 6b I included both repeating a relative year in school and acceleration in the same model. This is a piecemeal regression analysis, keeping in mind that for the majority of students who are neither accelerated nor repeat a yearin school, both these variables take on a value of zero. However, when both repeating a year in school and acceleration are included in the same model, the results are essentially the same as for considering each of them separately.

Summary

In summary, the negative RYiSE is almost unaffected by the inclusion of starting age, repeating a year in school or acceleration. Although some of these variables considered separately have small effects on M-ASC, consistent with the RYiSE (e.g., positive effects of starting school at a later age and repeating a RYiS), these effects are mostly explained by inclusion of relative year in school as a predictor variable. As expected, the inclusion of each of these variables had little if any effect on the negative effect of school-average achievement (the BFLPE). Although the effects of these three variables, and their juxtaposition with relative year in school are interesting, the sizes of these effects are sufficiently small to be of limited practical importance for the negative RYiSE. Consistent with my underlying premise about RYiSEs, these results suggest that relative year in school, rather than starting age, acceleration, or retention, is the critical variable.

Generalizability of BFLPE and RYiS Effects Over Student Background Variables (Research Question 2)

Here I evaluate how the negative BFLPEs and RYiSEs on M-ASC are related to student background characteristics: achievement, gender, age, immigrant status, home language, and socioeconomic status (SES). Although the relation of each of these variables with M-ASC is of interest in its own right, my primary focus is on their interactions with the negative effects of school-average achievement (the BFLPE) and the negative effects of relative year in school (RYiSE): that is, whether any of these background variables moderate the BFLPE or the RYiSE.

Achievement

Achievement is the potential moderator of the BFLPE that has attracted the most attention because of its obvious implications for policy/practice as well as theory (see earlier discussion): whether the effect of school-average achievement is consistently negative for students of different achievement levels, or perhaps the effect of school-average achievement is less negative or even positive for very bright students. Consistent with previous research (e.g., Huguet et al., 2009; Marsh, Kuyper et al., 2014; see reviews by Marsh, Seaton et al., 2008; Marsh & Seaton, in press), the interaction between individual and school-average achievement in the present investigation is small (−.059) relative to the size of the negative effect of school-average achievement (−.302) such that the direction of the BFLPE would still be negative even for students at the extremes of individual student achievement. Of particular importance from a policy/practice perspective, although the BFLPE is moderated to a small extent by individual student achievement, the interaction is negative rather than positive. Hence the negative effect of school-average achievement is slightly more negative for students with higher levels of achievement. Thus, high achievement slightly exacerbates rather than protects students from the negative consequences of the BFLPE.

Although not previously explored, moderation of the negative RYiSEs by individual student achievement also has similarly important theoretical and policy/practice implications: whether the negative effects of relative year in school are consistent across students of different achievement levels or, perhaps, the effects are less negative or even positive for very bright students. The negative RYiSE is not significantly moderated by achievement; the negative RYiSE generalizes across different levels of achievement.

Age

Although I have considered the effect of age as a potentially important moderating variable, main and interaction effects are consistently very small or nonsignificant. However, this is not surprising, given the highly truncated range of ages in the PISA data (i.e., all participants are 15-year-olds).

Gender

Boys have higher M-ASCs than girls (.114); this is consistent with one of the most robust gender differences in educational research. Although the interactions with gender are very small, gender interacts significantly with both RYiS (−.013) and the BFLPE (.019). Compared with girls, boys suffer slightly more negative RYiSEs but slightly less negative BFLPEs. However, these interactions are very small relative to the main effects of RYiS and the BFLPE, such that on average both boys and girls suffer negative RYiSEs and BFLPEs.

Socioeconomic status (SES)

For the PISA measure of global SES, there is a small positive effect of SES on M-ASC (.032). However, the interaction effects are very small (relative year in school: −.004; *SE* = .002; school-average achievement: −.008; *SE* = .003). Hence, negative effects of relative year in school and school-average achievement are slightly more negative for students from high-SES families. Thus, coming from a high-SES family slightly exacerbates rather than protects students from the negative consequences of the BFLPE and RYiSEs (see Supplemental Materials for analyses of different components of SES; family educational resources and possessions, parent education, and occupational status).

Immigration and language

The final two background characteristics are home language and immigrant status (see Supplemental Materials, Section 1, for further discussion of how these were defined). Immigrant and nonnative language students have slightly higher M-ASCs (Models 11 and 12, Table 4). Similar findings have been reported in the literature before, and have been described as the “immigrant paradox” (Fuligni, 1998; Sam & Berry, 2010); immigrant students show better psycho-social outcomes than their native-born counterparts. Thus, even though immigrant students tend to perform more poorly at school, they seem to have more positive attitudes toward **school**, higher aspirations, and more optimism for the future than their native-born counterparts (Fuligni, 1998; Portes & Rumbault, 2001; Sam & Berry, 2010). However, the interactions with these characteristics and the negative effects of RYiS are nonsignificant. Nonetheless, there are small but statistically significant interactions with the BFLPE. Immigrant and second-language students suffer slightly smaller BFLPEs than students who are not immigrants or whose home language is the native language.

Table 4

Moderating Effects: Models of BFLPE and RYiS Effects With Inclusion of Covariates and Their Interactions With BFLPE and RYiS

Effect	Model 7 Student Achievement β (*SE*)	Model 8 Age β (*SE*)	Model 9 Gender Male β (*SE*)	Model 10 SES Global β (*SE*)	Model 11 Home Language β (*SE*)	Model 12 Immigrant Status β (*SE*)
Fixed effects						
Student Linear ACH (SLA)	.552 (.019)	.549 (.019)	.528 (.019)	.543 (.019)	.555 (.019)	.555 (.019)
Student Quad ACH (SQA)	.126 (.005)	.108 (.006)	.102 (.006)	.108 (.006)	.106 (.006)	.105 (.006)
RYiS	−.096 (.009)	−.092 (.009)	−.080 (.008)	−.097 (.009)	−.093 (.009)	−.092 (.009)
School ACH (BFLPE)	−.302 (.023)	−.304 (.024)	−.286 (.023)	−.317 (.024)	−.300 (.023)	−.300 (.023)
Covariates (CV)		−.005 (.002)	.114 (.002)	.032 (.002)	.048 (.002)	.048 (.002)
CV × RYiS	−.005 (.003)	.002 (.002)	−.013 (.002)	−.004 (.002)	−.001 (.002)	−.000 (.002)
CV × BFLPE	−.059 (.006)	−.008 (.003)	.019 (.003)	−.008 (.003)	.017 (.003)	.012 (.003)
Random effects						
Level 3 intercept	.057 (.013)	.058 (.013)	.058 (.013)	.057 (.013)	.058 (.013)	.060 (.013)
SLA	.015 (.003)	.015 (.003)	.014 (.003)	.014 (.003)	.015 (.003)	.015 (.003)
SQA	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)	.001 (.000)
RYiS	.002 (.001)	.002 (.001)	.002 (.001)	.002 (.001)	.002 (.001)	.003 (.001)
BFLPE	.019 (.005)	.020 (.005)	.019 (.005)	.021 (.005)	.022 (.005)	.022 (.005)
Level 2 intercept	.029 (.001)	.029 (.001)	.027 (.001)	.029 (.001)	.029 (.001)	.029 (.001)
Level 1 intercept	.763 (.002)	.774 (.002)	.763 (.002)	.773 (.002)	.768 (.002)	.772 (.002)

Note. BFLPE = Big-Fish-Little-Pond-Effect; RYis = relative year in school; ACH = achievement. Covariates are: SLA = linear component of individual student achievement; Age = current age; Multiple components of socioeconomic status (SES) background (educational resources, highest level of parents' education, highest level of parents' occupational status, global); Starting Age = Age first started primary school (Year 1); Lang (language spoken at home, 1 = language of instruction and test, 0 = other); Immigrant (1 = first generation; 2 = second generation; 3 = nonimmigrant). Critical parameters, shaded in gray, are the effects of RYiS, school-average achievement (BFLPE), the covariates, and covariate interactions with RYiS and the BFLPE. All parameter estimates (βs) are significant at .05 level when they differ from zero by more than 2 *SEs*.

Moderating Effects: Models of BFLPE and RYiS Effects With Inclusion of Covariates and Their Interactions With BFLPE and RYiS

Summary

The results in this section demonstrate that both the negative RYiSEs and BFLPEs generalize very well over the student characteristics considered here. Although some of these student background characteristics did interact with either RYiS level or the BFLPE, the sizes of these interactions were consistently small, and much smaller than the negative RYiSEs and BFLPEs. Hence none of the moderations were sufficiently large to effect the negative direction of the negative RYiSEs and BFLPEs. In summary, the negative RYiSEs and BFLPEs generalized well over student differences in age, achievement levels, gender, home language, immigrant status, and SES.

Discussion

In the present investigation, theoretical models (e.g., frame of reference models such as social comparison theory) that are the basis of the widely studied BFLPE (the negative effect of school-average achievement on academic self-concept) are extended to encompass RYiSEs for M-ASC in PISA2003. In support of theoretical predictions, the empirical results demonstrate the negative effects of RYiSEs on M-ASC in addition to the already well-established negative effects of school-average achievement. These negative RYiSEs have a strong theoretical rationale based on the frame-of-reference effectsthat underpin the BFLPE. Empirically, these negative RYiSEs have broad generalizability across the 41 countries in PISA2003 and across many individual student characteristics considered as potential moderators of these effects. Although there is extensive support for theoretical predictions and empirical results based on the BFLPE, the present investigation is apparently the first to extend the theoretical model underpinning the BFLPE to incorporate RYiSEs. For this reason, I focus most of this discussion on the negative RYiSEs, which are an important new contribution. The rationale for the study is theoretical. However, the study also has important policy/practices implications about the age at which parents should start their children in school, as well as the effects of acceleration and retention. Furthermore, the study suggests that each of these different policy/practice strategies can be evaluated in relation to RYiS, thus, integrating these respective research literatures that are often considered in isolation.

Negative RYiSEs Cross-cultural generalizability

Our study is apparently the largest and probably the strongest cross-cultural study of the negative RYiSEs ever undertaken, based on nationally representative samples from 41 different countries (276,165 15-year-old students from 10,274 schools). The PISA2003 database provides a particularly powerful basis for evaluating the cross-cultural generalizability of results. Indeed, the major impediments to cross-cultural research (the lack of representative samples from many countries based on the same measures that are equally appropriate across cultures) are largely resolved through the use of PISA data. The results demonstrate that the negative RYiSEs generalized well over the 41 diverse countries, which varied substantially in relation to the structure of the school system. In particular, the negative RYiSEs on M-ASC were significantly negative in almost all of the countries, and were not significantly positive in any of the countries. Although there was country-to-country variation in these negative RYiSEs, the residual variance component representing this variation was small, and could be explained at least in part by a relatively few countries where all or nearly all 15-year-olds were in the same year in school. Hence, there is support for the cross-cultural generalizability of the negative RYiSEs.

Year in school: Integrating retention, acceleration and starting age

The theoretical rationale for negative RYiSEs, as posited here, is that the frame of reference depends on comparisons with classmates who are in the same year in school, whether these differences are because of school starting age, repeating grades, acceleration, changing countries, or any of a potentially large number of RYiS correlates—circumstances that might result in students being in a RYiS where they are systemically older or younger than their classmates (i.e., correlates of RYiS; see Figure 2). On this basis I proposed that RYiS would capture all or at least a substantial amount of the variance that can be explained by retention, acceleration, and starting age. The results showed good support for these predictions—at least in relation to the outcomes considered here—these RYiS correlates provide little or no additional information beyond RYiS. These results have potentially important implications, providing a link between research on starting age, retention, and acceleration, where there has been surprisingly little cross-fertilization research in these different areas. Indeed, I anticipate that stronger tests of this assumption will require stronger statistical models than have been used in past research into the effects of retention, acceleration, and starting age. Thus, for example, the Allen et al. (2009) meta-analysis of grade retention effects found that much of the effect attributed to retention was probably because of methodologically weak studies, while Wu et al. (2010) argued for evolving propensity matching approaches providing methodologically stronger tests of retention effects than statistical adjustment approaches.

Moderation by student level achievement

In some respects, the most interesting potential moderator is individual student achievement. Indeed, the simultaneous consideration both of individual and of school-average achievement in the same model highlights the importance of a multilevel perspective, in that what might be seen as the same variable has a substantially positive effect at the individual student level but a substantially negative effect at the group level. Thus they are in effect different variables, as is made clear in an appropriate multilevel perspective and associated analyses. Furthermore, the interaction between these two variables has been a particularly controversial issue in relation to the BFLPE. More specifically, some early research has suggested that the negative BFLPEs are larger for less able students—a finding that might support a more rigid tracking system based on more extensive measures of prior achievement. In contrast, as discussed earlier, the theoretical frame of reference model underpinning the BFLPE suggests that the negative effects of school-average achievement should not be moderated by individual student achievement. Over time, this controversy has been largely resolved, with increasing support for the theoretical prediction that the size of the BFLPE is relatively independent of individual student achievement—that the size of the interaction between individual and school-average achievement is consistently small, sometimes not statistically significant despite the very large sample sizes, and not even consistent in direction when it is statistically significant (see review by Marsh, Seaton et al., 2008; Marsh & Seaton, in press). In the present investigation, these theoretical predictions based on the BFLPE are extended to the negative RYiSEs. Consistently with the large body of BFLPE research, these new findings indicate that the negative RYiSEs are reasonably independent of individual achievement—the negative RYiSEs are of a similar size for brighter and weaker students.

Moderation by background variables

I also evaluated the extent to which the negative RYiSEs could be explained in terms of, or were moderated by, any of a set of background demographic variables (age, gender, socioeconomic status, home language, and immigration status). In the regression analyses, the background variables were controlled by simply including them into the model. Many of these background variables were related—albeit weakly in most cases—to M-ASC. Thus, for example, consistent with previous research, higher M-ASC is associated with being male, higher socioeconomic status, being an immigrant and a nonnative speaker (consistent with the “immigrant paradox” described earlier). Although these variables are interesting in their own right, the negative RYiSEs were almost unaffected by their inclusion. However, the critical question, in terms of the present investigation, is whether any of these background variables interacted with RYiS—whether the negative RYiSEs were moderated by the background variables. A few of the interaction effects were statistically significant, because of the huge sample size (e.g., RYiSEs are slightly larger for boys than girls). However, all these interaction effects were so tiny in size and in relation to the negative RYiSEs as to be substantively trivial. In particular, none of these interactions was sufficiently large to change the direction of the negative effects of the BFLPE or the RYiSEs.

Strengths, Limitations, and Directions for Further Research

A major strength of the current study is that it encompasses a very large, culturally diverse sample of students, nationally representative samples from 41 countries, a range of student background characteristics, and diverse outcome variables. The apparently idiosyncratic sampling design of the PISA data is sometimes seen as a weakness, or at least as a complication that needs to be addressed. However, this feature of PISA data is an important strength in evaluating RYiSEs. The PISA data are ideal for assessing the cross-cultural generalizability of the RYiSE, overcoming many of the traditional limitations of cross-cultural research by including nationally representative samples for a large number of different countries with diverse school systems. Indeed, the results provided good support for the cross-cultural generalizability of negative RYiSEs and BFLPEs; they were not significantly positive for any of the 41 countries and were significantly negative in most. However, there are also important limitations associated with the use of PISA data. Thus, for example, because this study was based on responses by 15-year-olds, there was no basis for extrapolating the results to other age groups—particularly for the negative RYiSEs, where there is no research literature at hand with which to evaluate age effects.

The focus of this study was specifically on the academic domain, but BFLPE-like effects have also been posited for physical outcomes in elite and nonelite sporting contexts (e.g., Marsh, Morin, & Parker, in press), where issues of acceleration, retention, and demotion might also be relevant. This is also relevant to suggestions by Marsh et al. (1995) that the negative effect of relative year in school is not domain-specific but generalizes to nonacademic (e.g., physical, social, and emotional) as well as academic domains of self-concept. This could not be tested in the present investigation, where the focus of PISA variables was on math constructs, but is a potentially important direction for further research.

Cross-sectional data

A major limitation of the present study is that it is based on a single wave of cross-sectional data, so that causality cannot be inferred. Educational studies routinely rely on such data where random assignment is not possible or would be ethically dubious. However, in relation to the BFLPE, this limitation should be viewed in the light of longitudinal, quasi-experimental, and true experimental studies of the BFLPE (see earlier discussion; also see Marsh, Seaton et al., 2008; Marsh & Seaton, in press), which demonstrate strong support for the construct validity of interpretations of enduring effects of the BFLPE (see also Marsh, 1991). The negative effect of school-average achievement may be confounded with factors such as individual differences in prior achievement, learning, family background, and school climate. However, as pointed out by Marsh, Hau, and Craven (2004), most of these potentially confounding factors are likely to be more positive in high-achievement schools, so that the direction of any such biases because of these uncontrolled sources of variation is likely to work against the BFLPE. In this respect, BFLPEs seem to be robust in relation to many potential biases.

Although it may be premature to argue that the same rationale applies to negative RYiSEs based on cross-sectional studies, a similar logic seems to apply. Thus, for retained students, uncontrolled, pre-existing differences leading to retention are likely to be negatively related to most outcomes. Because this would result in negatively biased estimates of retention effects, statistical adjustment procedures are likely to underadjust for circumstances leading to retention. Indeed, this pattern of results was identified in the Allen et al. (2009) meta-analysis, which showed that methodologically weaker studies resulted in negatively biased estimates of retention effects, compared with methodologically stronger studies with better controls for pre-existing differences. Similarly, for accelerated students, uncontrolled, pre-existing differences leading to acceleration are likely to produce positively biased estimates. In each case, the direction of bias is likely to work against the negative RYiSEs, suggesting that stronger designs would be likely to result in even more negative RYiSEs. Furthermore, the results of the present investigation demonstrate that a variety of potential confounding variables had little or no influence on either the BFLPE or the RYiSE. In addition, although the present investigation is clearly cross-sectional in relation to the outcome variables considered, it is at least quasi-longitudinal in relation to RYiS and correlates of RYiS: these were temporally before outcomes that served as dependent variables in the present investigation. Although there is an abundance of research in support of the construct validity of BFLPE interpretations (e.g., Marsh, Seaton et al., 2008) in relation to longitudinal, quasi-experimental and true random assignment designs, more research is needed to establish the robustness of negative RYiSEs in relation to stronger experimental designs. Nevertheless, these are issues that can be addressed by future studies of the RYiSEs, perhaps modeled on the research used to evaluate the robustness of the BFLPE.

In summary, there are also important limitations in the interpretation of effects based on cross-sectional data. In particular, because this is apparently the first large-scale, rigorous study of negative RYiSEs, it is premature to claim support for their universality. However, if further research replicates the robustness of the RYiSE, it might be appropriate to consider it a pan-human phenomenon, as has been suggested for the BFLPE (e.g., Seaton et al., 2010).

Generalizability of Negative Effects of School-Average Achievement and RYiSEs to Other Constructs and Underlying Processes

In the present investigation I focused mainly on the negative effects of school-average achievement (the BFLPE) and the negative effects of relative year in school (RYiSEs) on M-ASC that has been the basis of most previous BFLPE research. However, to better understand the generalizability, nature, and processes underlying the negative effects of school-average achievement and relative year in school it is useful to evaluate how these variables are related to the other constructs. Although beyond the scope of the present investigation, in supplemental analyses I tested the extent to which the negative effects of school-average achievement and RYiS generalize to other math-related psychosocial variables available in the PISA database: interest/enjoyment in maths, instrumental motivation in maths, math anxiety, math self-concept, different learning strategies (Memorisation/rehearsal, Elaboration, Control), and orientation preferences (Competitive, Cooperative). (See Supplemental Materials, Section 1, for a detailed definition and operationalization of each of these variables, and Section 6 for presentation of the analyses and results.) As anticipated, the negative effects of school-average achievement (in support of previous BFLPE research) and negative RYiSEs (where there is not much previous research) were both systematically smaller than those observed with M-ASC. However, in support of the generalizability of the negative effects of school-average achievement and relative year in school, the effects were largely negative and none were significantly positive for any of these outcomes.

Following from these supplemental analyses it would be useful to know the extent to which these or other variables mediated the negative effects of school-average achievement or RYiSEs. Thus, for example, if it could be established that either the BFLPE or the negative RYiSEs were completely mediated by a critical mediating variable, it would not only provide a better understanding of the effects but might also elicit policy implications on how to undermine the negative consequences of these effects. However, because the negative effects of school-average achievement and relative year in school are systematically larger for M-ASC than for other constructs, it is clear that the negative BFLPEs and RYiSEs cannot be completely explained by any of these other variables.

Particularly with evolving sophistication in tests of mediation (e.g., Nagengast & Marsh, 2012; Preacher, Zyphur, & Zhang, 2010), it is increasingly easy to test mediation hypotheses. However, it remains extremely difficult to test the validity of a mediation interpretation, because of the strong assumptions about causal ordering (i.e., independent variable → Mediator → dependent variable) that are implicit and typically untested or even untestable in most mediation analyses based on cross-sectional data studies like the present investigation. Indeed, the causal assumptions underlying mediation tests are difficult to validate in longitudinal studies and even true experimental studies with random assignment to the intervention but not the mediating variable.

It is, nevertheless, heuristic for future research to speculate on what some of the processes underlying the negative effects of relative year in might be and juxtapose these with existing BFLPE research. Thus, for example, Marsh and Seaton (in press)evaluated support for suggested strategies to reduce the negative BFLPEs that might also generalize to negative RYiSEs:

- Focus on individual improvement, achieving personal bests, and mastery of new skills rather than doing better than classmates or being the best student their class.
- Encourage students to pursue their own goals of particular interest to them, reducing social comparison.
- Avoid highly competitive environments that encourage social comparison processes underlying BFLPEs.
- Providing individual feedback in relation to criterion reference standards and personal improvement over time rather than comparisons based on the performances of other students.

However, Marsh and Seaton (in press) concluded that was either insufficient research in relation to the BFLPE to support these heuristic suggestions or, where there was relevant research, it did not support the suggested strategies. Nevertheless, they also suggested that a particularly useful direction for further research is to consider intervention studies designed to alter motivational climates so as to undermine social comparison processes that underlie the BFLPE. Although these issues have been considered in BFLPE research—even if further research is needed—there is apparently little or no research to test these strategies in relation to RYiSEs.

Summary and Implications

The results of the present investigation have important implications for theoretical models of social comparison processes, for frame-of-reference effects, and also for educational policymakers worldwide. The present investigation has extended theory by showing two very different frame-of-reference effects to be evident in numerous, culturally diverse countries, and to be consistent across achievement levels; thus, earning consideration as pan-human theoretical models. In regard to educational policy, in many countries around the world, high-achievement students are increasingly being taught in academically selective schools, while the collected body of BFLPE research reviewed here—as well as the present investigation—suggests that these may not be the optimal environment for such students, at least in terms of the diverse set of psycho-social variables considered here and in other BFLPE research.

Similarly, educational policymakers in different countries use diverse strategies in relation to school starting age, repeating grades, and acceleration, apparently without fully understanding the implications of these policy practices in relation to academic self-concept, motivation, and a range of affective variables that have long-term implications for academic choice and accomplishments. Particularly since the results of the present investigation are contrary to at least some of the accepted wisdom in relation to acceleration and retention used by parents and schools, there is need for further research to more fully evaluate the generalizability and construct validity of interpretations offered here. However, results from the present investigation clearly call into question any simplistic conclusions that acceleration is “good” and retention is “bad.”

In summary, the major new contributions of the present investigation are to provide tests of and support for:

- New theoretical predictions based on the frame of reference theory underpinning the BFLPE, as applied to the effects of RYiS on academic self-concept. The theoretical basis of these predictions appears to be entirely new and has not been previously considered in the research literatures on school starting age, retention (repeating a grade), or acceleration (skipping a year in school);

- New theoretical predictions that the effects of retention, acceleration, and starting age can all be encapsulated in a single variable, relative RYiS;
- Challenges to prevailing beliefs that retention has negative effects and acceleration has positive effects on academic self-concept;
- The juxtaposition of RYiSEs and BFLPEs, demonstrating that these are relatively independent effects, even though they both have a similar theoretical basis.

References

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary school grades*. New York, NY: Cambridge University Press.

Allen, C., Chen, Q., Willson, V., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis*, *31*, 480–499. 10.3102/0162373709352239

Alwin, D. F., & Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education*, *50*, 259–273. 10.2307/2112499

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

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

Boyd, L., & Iverson, G. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth.

Chen, S., Yeh, Y., Hwang, F., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort-multioccasion study. *Learning and Individual Differences*, *23*, 172–178. 10.1016/j.lindif.2012.07.021

Coleman, J. M., & Fults, B. A. (1985). Special class placement, level of intelligence, and the self-concepts of gifted children: A social comparison perspective. *Remedial and Special Education*, *6*, 8–10. 10.1177/074193258500600103

Davis, J. A. (1966). The campus as a frog pond: An application of theory of relative deprivation to career decisions for college men. *American Journal of Sociology*, *72*, 17–31. 10.1086/224257

Diener, E. (2000). Subjective well-being—The science of happiness and a proposal for a national index. *American Psychologist*, *55*, 34–43. 10.1037/0003-066X.55.1.34

Diener, E., & Fujita, F. (1997). Social comparison and subjective well-being. In B. P.Buunk & F. X.Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–358). Mahwah, NJ: Erlbaum.

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

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review, 43*, 557–572. 10.2307/2094779

Fuligni, A. (1998). The adjustment of children from immigrant families. *Current Directions in Psychological Science*, *7*, 99–103. 10.1111/1467-8721.ep10774731

Gladwell, M. (2013). *David and Goliath: Underdogs, misfits, and the art of battling giants*. Boston, MA: Little, Brown & Company.

Goldstein, H. (2003). *Multilevel statistical models*. London: Edward Arnold.

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity*, *3*, 53–68. 10.1080/13576500342000040

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*, 124–136. 10.1037/0022-0663.95.1.124

Hattie, J. A. (2012). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon, England: Routledge.

Helson, H. (1964). *Adaptation-level theory*. New York, NY: Harper & Row.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., . . .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*, 156–170. 10.1037/a0015558

Hyman, H. (1942). The psychology of subjective status. *Psychological Bulletin, 39*, 473–474.

Im, M. H., Hughes, J. N., Kwok, O., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology*, *51*, 349–365. 10.1016/j.jsp.2013.01.004

James, W. (1963). *The principles of psychology* (Vol. *2*). New York, NY: Holt. (Original work published 1890)

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *SchoolPsychology Review*, *30*, 420–437.

Jimerson, S. R., & Brown, J. A. (2013). Grade retention. In J. A.Hattie & E. M.Anderman (Eds.), *International guide to student achievement* (pp. 42–44). New York, NY: Routledge.

Jonkmann, K., Becker, M., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2012). Personality traits moderate the Big-Fish–Little-Pond effect of academic self-concept. *Learning and Individual Differences*, *22*, 736–746. 10.1016/j.lindif.2012.07.020

Kretschmann, J., Vock, M., & Lüdtke, O. (2014). Acceleration in elementary school: Using propensity score matching to estimate the effects on academic achievement. [Advance online publication]. *Journal of Educational Psychology, 106,* 1080–1095. 10.1037/a0036631

Kulik, J. A., & Kulik, C. C. (1984). Effects of accelerated instruction on students. *Review of Educational Research*, *54*, 409–425. 10.3102/00346543054003409

Kulik, J. A., & Kulik, C.-L. C. (2004). Meta-analytic findings on grouping programs. In L. E.Brody (Ed.), Essential reading in gifted education. *Grouping and acceleration practices in gifted education* (pp. 105–114). Thousand Oaks, CA: Corwin Press.

Lüdtke, O., Köller, O., Marsh, H., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology, 30,* 263–285. 10.1016/j.cedpsych.2004.10.002

Marsh, H. W. (1974). *Judgmental anchoring: Stimulus and response variables*. Unpublished doctoral dissertation, University of California, Los Angeles.

Marsh, H. W. (1983). Neutral-point anchoring of personality-trait words. *The American Journal of Psychology*, *96,* 513–526. 10.2307/1422572

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education, 28,* 165–181.

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

Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2,* 77–172. 10.1007/BF01322177

Marsh, H. W. (1991). Failure of high ability schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal*, *28,* 445–480. 10.3102/00028312028002445

Marsh, H. W. (2005). Big fish little pond effect on academic self-concept. *German Journal of Educational Psychology*, *19*, 119–128.

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

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of US and Saudi Arabian TIMSS responses. *Journal of Cross-CulturalPsychology*, *45*, 777–804. 10.1177/0022022113519858

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. A. (2015). 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* (on-line pre-publication version).

Marsh, H. W., Chessor, D., Craven, R., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal*, *32*, 285–319. 10.3102/00028312032002285

Marsh, H. W., & Craven, R. (2002). The pivotal role of frames of reference in academic self-concept formation: The big fish little pond effect. In F.Pajares & T.Urdan (Eds.), *Adolescence and education* (Vol. *II*, pp. 83–123). Charlotte, NC: Information Age Publishing.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*, 133–163. 10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K. (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*, 364–376. 10.1037/0003-066X.58.5.364

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of EducationalPsychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*, 311–360. 10.1207/s15327574ijt0604_1

Marsh, H. W., Hau, K., & Craven, R. G. (2004). The big-fish-little-pond effect stands up to scrutiny. *American Psychologist*, *59*, 269–271. 10.1037/0003-066X.59.4.269

Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big fish little pond effect on academic self-concept. *American Educational Research Journal*, *38*, 321–350. 10.3102/00028312038002321

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*, 337–349. 10.1037/0022-3514.78.2.337

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. 10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *EducationalPsychologist*, *47*, 106–124. 10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthen, 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*, 764–802. 10.1080/00273170903333665

Marsh, H. W., Morin, A. J. S., & Parker, P. D. (in press). Physical self-concept changes in a selective sport high school: A longitudinal cohort-sequence analysis of the big-fish-little-pond effect. *Journal of Sport & Exercise Psychology*.

Marsh, H. W., & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educationaloutcomes: Direct and indirect effects of the big-fish-little-pond effect. *Zeitschrift für Pädagogische Psychologie/German*. Journal of Educational Psychology, *24*, 51–72.

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*, 213–231. 10.1037/0022-3514.47.1.213

Marsh, H. W., & Seaton, M. (in press). The big-fish-little pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. In A.Elliot (Ed.), *Advances in motivation science*. Oxford, UK: Elsevier.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational PsychologyReview*, *20*, 319–350. 10.1007/s10648-008-9075-6

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

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*, 510–524. 10.1037/0022-0663.100.3.510

Marsh, H. W., Walker, R., & Debus, R. (1991). Subject-specific components of academic self-concept and self-efficacy. *Contemporary Educational Psychology*, *16*, 331–345. 10.1016/0361-476X(91)90013-B

Meyer, J. W. (1970). High school effects on college intentions. *American Journal of Sociology*, *76*, 59–70. 10.1086/224906

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology*, *16*, 148–156. 10.1037/h0029862

Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of math and reading achievement in low achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology*, *104*, 603–621. 10.1037/a0027571

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*, 1033–1053. 10.1037/a0027697

Nagengast, B., & Marsh, H. W. (2013). Motivation and engagement in science around the globe: Testing measurement invariance with multigroup SEMs across 57 countries using PISA 2006. In L.Rutkowski, M.von Davier, & D.Rutkowski (Eds.), *A handbook of international large-scale assessment data analysis* (pp. 317–344). London: Chapman & Hall, CRC Press.

Organisation for Economic Cooperation and Development. (2005a). *PISA 2003 data analysis manual: SPSS users*. Paris: Author.

Organisation for Economic Cooperation and Development. (2005b). *PISA 2003 Tech. Rep. No*. Paris: Author.

Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Mahwah, NJ: Erlbaum.

Parducci, A., Perrett, D., & Marsh, H. W. (1969). Assimilation and contrast as range frequency effects of anchors. *Journal of Experimental Psychology*, *81*, 281–288. 10.1037/h0027741

Pinxten, M., de Fraine, B., van Damme, J., & D'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. *British Journal of Educational Psychology*, *80*, 689–709. 10.1348/000709910X493071

Portes, A., & Rumbault, R. (2001). *Legacies: The story of the second generation*. Berkeley, CA: University of California Press.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233. 10.1037/a0020141

Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2004). *A user's guide to MLwiN–Version 2.0*. Bristol, TN: University of Bristol.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Reynolds, A. J. (1992). Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation and Policy Analysis*, *14*, 101–121. 10.3102/01623737014002101

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, *31*, 729–759. 10.3102/00028312031004729

Roderick, M., & Engel, M. (2001). The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing. *Educational Evaluation and Policy Analysis*, *23,* 197–227. 10.3102/01623737023003197

Rogers, S. (1941). The anchoring of absolute judgments. *Archives de Psychologie*, *37*, 1–42.

Sam, D. L., & Berry, J. W. (2010). Acculturation: When individuals and groups of different cultural background meet. *Perspectives on Psychological Science*, *5*, 472–481. 10.1177/1745691610373075

Schulz, W. (2005). Scaling procedures and construct validation of context questionnaire data. In R.Adams (Ed.), *PISA 2003 Tech. Rep. No.*Paris: OECD Publications.
Schwarzer, R., Lange, B., & Jerusalem, M. (1982). Selbstkonzeptentwicklung nach einem Bezugsgruppenwechsel [Self-concept development after a reference-group change]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *14*, 125–140.

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*, 403–419. 10.1037/a0013838

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-Fish-Little-Pond-Effect: Generalizability and moderation—Two sides of the same coin. *American Educational Research Journal*, *47*, 390–433. 10.3102/0002831209350493

Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., . . .Suls, J. (2008). In search of the big-fish: Investigating the coexistence of the Big-Fish-Little-Pond effect with the positive effects of upward comparison. *British Journal of Social Psychology*, *47*, 73–103. 10.1348/014466607X202309

Seaton, M., Marsh, H. W., & Parker, P. D. (2013). A tale of many countries: A review of Big-Fish-Little-Pond Effect research in different cultural contexts. In G. A.Liem & A. B. I.Bernardo (Eds.), *Advancing cross-cultural perspectives on educationalpsychology: A festschrift for Dennis McInerney* (pp. 211–227). Greenwich, CT: Information Age Press.

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. *American Psychologist*, *53,* 1101–1110. 10.1037/0003-066X.53.10.1101

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 5–14.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. *Review of EducationalResearch*, *46*, 407–441.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, No. *187*.

Sherif, M., & Sherif, C. W. (1969). *Social psychology*. New York, NY: Harper & Row.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M. (1949). *The American soldier: Adjustments during army life* (Vol. *1*). Princeton, NJ: Princeton University Press.

Suls, J. M. (1977). Social comparison theory and research: An overview from 1954. In J. M.Suls & R. L.Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 1–20). Washington, DC: Hemisphere.

Suls, J., & Wheeler, L. (2000). A selective history of classic and neo-social comparison theory. In J.Suls & L.Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 3–19). New York, NY: Kluwer Academic/Plenum Press. 10.1007/978-1-4615-4237-7_1

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of EducationalPsychology*, *98*, 788–806. 10.1037/0022-0663.98.4.788

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, *101*, 853–866. 10.1037/a0016306
Upshaw, H. S. (1969). In L.Berkowitz (Ed.), The personal reference scale: An approach to social judgment. In L.Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. *4*, pp. 315–370). New York, NY: Academic.

Volkman, J. (1951). Scales of judgment and their implications for social psychology. In J. H.Loher & M.Sherif (Eds.), *Social psychology at the crossroads* (pp. 273–294). New York, NY: Harper.

Wedell, D. H., & Parducci, A. (2000). Social comparison: Lessons from basic research on judgment. In J.Suls & L.Wheeler (Eds.), *Handbook of social comparison: Theory and research. The Plenum Press series in social/clinical psychology* (pp. 223–252). Dordrecht, Netherlands: Kluwer Academic Publishers. 10.1007/978-1-4615-4237-7_12

Wu, W., West, S. G., & Hughes, J. N. (2008). Effect of retention in first grade on children's achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology*, *100*, 727–740. 10.1037/a0013098

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology*, *102*, 135–152. 10.1037/a0016664