Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects : Pan-human cross-cultural generalizability over 68 countries

Marsh, Herbert W., Parker, Philip D., Guo, Jiesi, Pekrun, Reinhard and Basarkod, Geetanjali

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

This is an Accepted Manuscript of an article published European Journal of Personality, Eur. J. Pers. 34: 180–202 (2020) Published online 27 January 2020 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/per.223

Marsh, H. W., Parker, P.D., Guo, J. Pekrun, R. and BASARKOD1, G, (2020). Psychological Comparison Processes and Self–Concept in Relation to Five Distinct Frame–Of–Reference Effects: Pan–Human Cross–Cultural Generalizability over 68 Countries. *European Journal of Personality*., *34*(2), 180–202. https://doi.org/10.1002/per.2232

Running Head: Generalizability Of Psychological Comparison Processes

Psychological Comparison Processes and Self-Concept in Relation to Five Distinct Frame-of-Reference Effects: A Pan-Human Cross-Cultural Generalizability Over 68 Countries

Herbert W. Marsh [a,b], Philip D. Parker [a], Jiesi Guo [a], Reinhard Pekrun [c,a], Geetanjali Basarkod [a]
[a] Institute for Positive Psychology and Education, Australian Catholic University; [b] Department of Education, University of Oxford; [c] Department of Psychology, University of Munich

Revision submitted: 24 June, 2019
Revision submitted: 13 September, 2019

Corresponding Author: Herbert W. Marsh, Institute for Positive Psychology and Education (IPPE), Level 10, 33 Berry Street, North Sydney 2060 , Herb.Marsh@acu.edu.au; +61 2 9701 4626

Philip D. Parker, IPPE, North Sydney Campus, Level 9, 33 Berry Street, North Sydney 2060. PO Box 968, North Sydney NSW 2059. E: Philip.Parker@acu.edu.au

Jiesi Guo, IPPE, North Sydney Campus, Level 9, 33 Berry Street, North Sydney 2060. PO Box 968, North Sydney NSW 2059. E: Jiesi.Guo@acu.edu.au

Reinhard Pekrun, IPPE, North Sydney Campus, Level 9, 33 Berry Street, North Sydney 2060. PO Box 968, North Sydney NSW 2059. E: pekrun@lmu.de

Geetanjali Basarkod, IPPE, North Sydney Campus, Level 9, 33 Berry Street, North Sydney 2060. PO Box 968, North Sydney NSW 2059. E: Geetanjali.Basarkod@acu.edu.au

Author Note

The author would like to acknowledge Theresa Dicke for her helpful comments in relation to this article, as well as the many colleagues who have contributed to the research leading up to this research. This research was supported in part by a grant from the Australian Research Council to H. Marsh (DP130102713). There are no conflicts of interest related to this article.

Please address correspondence to Herbert W. Marsh, Institute for Positive Psychology and Education, Level 10, 33 Berry Street, North Sydney 2060, E-mail: Herb.Marsh@acu.edu.au

Acknowledgements The author would like to acknowledge Theresa Dicke for her helpful comments in relation to this article, as well as the many colleagues who have contributed to the research leading up to this research. This research was supported in part by a grant from the Australian Research Council to H. Marsh (DP130102713). There are no conflicts of interest related to this article.

Abstract

The concept of self is central to personhood, but personality research has largely ignored the relevance of recent advances in self-concept theory: multidimensionality of self-concept (focusing instead on self-esteem, an implicit unidimensional approach), domain specificity (generalizability of trait manifestations over different domains), and multilevel perspectives in which social-cognitive processes and contextual effects drive self-perceptions at different levels (individual, school, country) aligned to Bronfenbrenner's (1979) ecological model. Here we provide theoretical and empirical support for psychological comparison processes that influence self-perceptions and their relation to distal outcomes. Our meta-theoretical integration of social and dimensional comparison theories demonstrates five seemingly paradoxical frame-of-reference and contextual effects in self-concept formation that occur at different levels (68 countries/regions, 18,292 schools, 485,490 individual 15-year-old students). Consistent with dimensional comparison theory, the effects on math self-concept (MSC) were positive for math achievement, but negative for verbal achievement. Consistent with social comparison theory, the effects on MSC were negative for school-average math achievement (big-fish-little-pond effect, BFLPE), country-average achievement (paradoxical cross-cultural effect), and being young relative to year in school, but positive for school-average verbal achievement (BFLPE-compensatory effect). We demonstrate cross-cultural generalizability/universality of support for predictions over 68 countries/regions and discuss implications for relevance to personality research.

Keywords: Academic self-concept; social comparison theory and frame-of-reference effects; cross-cultural self-concept paradox; big-fish-little-pond effect; dimensional comparison theory

The overarching aim of this article is to develop an integrative meta-theory and statistical model of the formation of self-concept, test its cross-cultural generalizability, and outline its relevance to personality research. The key features of self-concept research that have not been well-integrated into personality research include the following:

- The domain specificity of self-concept based on a multidimensional model of self-concept rather than an implicit unidimensional approach in which self-concept is represented by a single, global factor (self-esteem, the higher-order factor in multidimensional, hierarchical models of selfconcept).
- The critical role of psychological comparison processes (social and dimensional) in the formation of particular self-perceptions that drive important outcomes.
- Treating self-reported personality as true self-perceptions (as is the case in self-concept research); this would facilitate the study of psychological comparison processes that mediate the effect of personality on self-perceptions, and the role of self-perceptions in mediating the effects of personality traits on subsequent outcomes.
- Domain specificity, the extent to which manifestations of particular traits generalize over different domains (e.g., are neurotic individuals equally neurotic in relation to social, family, sporting, and academic domains?) and the extent to which domain-specific traits are more strongly related to domain-specific outcomes (refer to subsequent discussion of the ‘specificity-matching principle’).
- Multilevel perspectives, as in the classic Bronfenbrenner (1979) ecological model of human development in which context at different levels influences characteristic adaptations like self-concept.
- Cross-cultural tests of the generalizability of key findings in personality research that are based substantially on studies of middle-class participants from Western countries.

Setting the stage, we begin with an overview of the importance of the self-concept construct, limitations in the incorporation of self into mainstream personality research, and our meta-theoretical model of the construct. We then posit a priori predictions on the basis of this meta-theoretical model and test the cross-cultural generalizability of these predictions using Programme for International Student Assessment (PISA) 2012 data. In conclusion, we discuss the implications of this integrative model to self-concept theory, and its relevance to personality research.

The importance of the self-concept construct

The construct of self has a long and distinguished history, tracing back as far as Socrates and Plato. In recent times, positive self-beliefs have become a key psychological construct in many areas of psychology and the social sciences more generally:

> ... a basic psychological need that has a pervasive impact on daily life, cognition and behavior, across age and culture . . . an ideal cornerstone on which to rest the achievement motivation literature but also a foundational building block for any theory of personality, development and well-being. (Elliot & Dweck, 2005, p. 8)

Self-concept and positive self-beliefs are considered a facilitator of success in social and emotional situations (Harter, 2012; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017; Pekrun, Murayama, Marsh, Goetz, & Frenzel, 2019), in school settings (Chen, Yeh, Hwang, & Lin, 2013; Marsh & Craven, 2006; Marsh & Yeung, 1997), and in everyday life (Eccles, 2009; Elliot & Dweck, 2005). Furthermore, as emphasized by Bandura (1986, 2006), Harter (2012), Marsh (2007), and others, positive self-belief stands alone as an important outcome, in addition to facilitating other desirable outcomes. Indeed, whereas negative self-beliefs and a lack of confidence may undermine pursuit of life's important challenges, positive self-beliefs give people the confidence to embark on new adventures that make good things happen.

The role of self in personality research

The construct of self—including self-beliefs, self-structures, self-related frame-of-reference effects, self-efficacy, and self-processes—has a somewhat ambiguous history in personality research and theory. This ambiguity is likely due to the tension between the assumed importance of self to personality and the difficulty of incorporating the various components of self into personality theory. Early personality theory emphasized the development of self (e.g., Allport, 1937, 1968; Lewin, 1935; also James, 1890). More recently, Greenwald (1988, p. 30) emphasized the central importance of self to personality theory, noting that it is 'a major (perhaps the major) structure of personality' and that 'differences between persons in their manner of, and effectiveness in establishing self-worth are fundamental to personality' (p. 37). Baumeister (1999), defining self-concept as the 'individual's belief about himself or herself, including the person's attributes and who and what the self is' (p. 681), argued that the 'self is not really a single topic at all, but rather an aggregate of loosely related subtopics' (p. 681). However, nearly all the manifestations of the self imply the capacity for self-reflection that is central to what it means to have a self, that distinguishes humans from most animals, and that is critical to understanding human behaviour. Nevertheless, the study of self is diffused across different sub-disciplinesof psychology, and integrative, cross-disciplinary research is rare.

Mischel and Morf (2003, p. 25; also Leary & Tangney, 2003) argued that historical accident and tradition 'have landed the self more or less in the province of social psychology and particularly social cognition, whereas the person mostly divided from social contexts became the domain of personality psychology'. Therefore, perhaps, personality researchers have often failed to integrate relevant selfprocesses into the personality processing system and the dominant research on personality traits. As such, a separation might be dysfunctional for both self and personality theory, and thus, Mischel and Morf called for an integration of these fields to build a more all-embracing, cross-disciplinary science. For example, studies of academic self-concept (ASC) have been considered largely in educational psychology and related disciplines. Yet the psychological comparison processes that underlie ASC formation and its relation to academic achievement and long-term outcomes are also at the heart of Festinger's (1954) social comparison theory and related social comparison processes that are relevant to personality theory (Huguet et al., 2009; also Hair & Graziano, 2003).

Similarly, McAdams (1996) emphasized that personality psychology has struggled to provide a conceptual framework that orients human individuality and personality within social and cultural contexts. Reiterating these issues, McAdams and Pals (2006) argue that personality research has failed in its mission to provide an integrated framework for understanding the whole person. Building on the classical William James (1890) distinction between 'I' and 'me' features of personality, McAdams (1996) notes that the 'me' reflects the content of the self-concept whereas the 'I' reflects the psychological processes used to construct the self-concept. Also building on William James' distinction is the multidimenational, hierarchical model of self-concept (Marsh & Shavelson, 1985; Shavelson, Hubner, & Stanton, 1976) with self-esteem at the apex and increasingly domain-specific components of selfconcept (e.g., academic, physical, social, and emotional) at lower levels. Hence, the multidimensional, hierarchical model of self-concept represents the 'me', whereas the psychological comparison processes represent part of the 'I'. For McAdams, personality traits such as the Big Five are part of the selfconcept to the extent that individuals consider these traits to be a part of who they are. Underscoring this difficulty for personality research to integrate social contexts, the Ozer and Benet-Martínez's (2006; also Baumert et al., 2017) review argues that for many, self-concept is an integral part of personality but that self-concept can also be seen as an outcome that is influenced by personality traits as well as life experience and context.

Historically, personality researchers who differentiated personality from self-concept viewed self-concept as more fluid than core personality traits (Marsh, 2008). Despite their robustness, the Big Five factors, a popular model of personality, cannot account for the complexity of personality as a whole (McCrae & Costa, 1996; Waller & Ben-Porath, 1987). Thus, Costa and McCrae (1992; also McAdams, 1996) distinguished between basic tendencies such as the highly stable Big Five factors, and surface characteristics that include more malleable aspects of personality (Shafer, 2000). Here, relations between personality factors and outcomes are mediated by these more malleable constructs, such as selfconcept and self-esteem in particular (the global, higher-order component of self-concept). For example, Parker, Martin, and Marsh (2008) found support for a process model of personality in which the relations between Big Five factors and life satisfaction were substantially mediated by multiple dimensions of self-concept. Similarly, in educational settings, Graziano, Jensen-Campbell, and Finch (1997; also Hair & Graziano, 2003) provided support for a process model in which specific components of self-concept act as mediators between stable personality traits and academic outcomes. We also note, however, that longitudinal studies show that specific domains of self-concept can be as stable over time as traditional personality traits (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006; also Marsh, 2008). Furthermore, in an evaluation of age-related differences in Big Five personality traits for a large representative sample (ages 15–99), Marsh, Nagengast, and Morin (2013) found what they referred to as the La Dolce Vita effect; in later years, individuals become happier (more agreeable and less neurotic), more self-content and self-centred (less extroverted and open), more laid back and satisfied with what they have (less conscientious, open, outgoing and extroverted), and less preoccupied with productivity.

In both self-concept and personality research, the key constructs are considered to be multidimensional. However, Marsh, Trautwein, et al. (2006; also Marsh, 2008) noted that in personality research, when self-concept is considered, it is often treated as a unidimensional construct based on global self-esteem measures. Although the multidimensional, hierarchical models of self-concept have been well-known since at least the 1980s when the Shavelson and Marsh/Shavelson models were published (Marsh & Shavelson, 1985;Shavelson et al., 1976), this perspective has rarely been integrated into personality research (refer to Marsh, Trautwein, et al., 2006, for an exception).

In a study examining the link between self-concept, personality, well-being, and academic outcomes, Marsh, Trautwein, et al. (2006) found that the personality and wellbeing factors did not contribute to the prediction of academic outcomes beyond what could be explained by ASC factors. Conversely, self-esteem by itself explained almost none of the variance in the personality or well-being factors. The results are consistent with the specificity-matching principle (Epstein, 1979; Swann, Chang-Schneider, & McClarty, 2007) that posits that specific dimensions of a construct should correlate more with domain-specific outcomes in the matching domain, rather than with broad outcomes. Swann et al. (2007) noted that modest relations between self-esteem and specific outcomes were to be expected in relation to the specificity-matching principle but emphasized the more impressive relations between specific components of selfconcept and important social outcomes. Thus, they concluded that 'people's self-views do matter, and the task of future researchers is to determine how, when, and with what consequences' (p. 92).

Taken together, these arguments, along with adherence to the specificity-matching principle, indicate the need for a multidimensional perspective on self-concept in personality research. In addition, the focus should be on those specific components of self-concept that are most logically related to the aim of the research (as well as on the dynamic processes that are important in the formation of self-perceptions in specific domains) rather than relying on global self-esteem measures. The present research explores some implications of this proposition by evaluating the role of psychological comparison processes in the formation of ASC in a large, cross-cultural sample.

Psychological comparison processes and frame-of-reference effects

Ubiquitous psychological comparison processes

Self-concepts are based on objective accomplishments evaluated in relation to frames of reference. James (1890) noted that 'we have the paradox of a man shamed to death because he is only the second pugilist or the second oarsman in the world' (p. 310). This perspective is central to the social comparison theory (Festinger, 1954) in psychology and has also been influential in sociology and economics. Extending this theoretical work, Marsh (1984; also Fang et al., 2018; Marsh & Parker, 1984; Marsh et al., 2008) proposed the big-fish-little-pond effect (BFLPE), in which students show lower ASCs when placed in a context of high-ability students but higher ASCs when placed in a context of low-ability students. Increasingly sophisticated theoretical and statistical models of ASC development, based in part on this BFLPE, have demonstrated that frame-of-reference effects occur at the school and country levels, as well as the individual-student level (Marsh, 2016; Marsh et al., 2018; Marsh & Seaton, 2015).

Comparative processes, or frames of reference, serve an important purpose in forming self-perceptions. Importantly, multiple frames of reference can be used to assess one′s accomplishments, and this, in turn, forms the basis of self-perceptions (Skaalvik & Skaalvik, 2002). In the broader psychological literature, the two best-known psychological comparison processes are social and temporal comparisons (Albert, 1977; Festinger, 1954; Möller & Marsh, 2013; Möller, Pohlmann, Köller, & Marsh, 2009); self-concept formation is based in part on comparisons between current and past accomplishments (temporal comparisons) and within a given social context (e.g., classmates in one′s class or school) and the comparisons between our own accomplishments and those of others (social comparisons). Marsh (1986; also Möller & Marsh, 2013) has also argued for the importance of a third psychological comparison process: dimensional comparisons that involve contrasting one′s own achievement in different areas (e.g.,, math achievement and verbal achievement, and relative performances in different events in a triathlon).

Support for these comparison processes has been shown not only in psychological research but also in other fields like economics (e.g., relative rather than absolute income impacts well-being; Clark, 2018). Thus, economist Frank (2012) posited an evolutionary (Darwinian) perspective in which the tendency to compare ourselves to immediate others is a fundamental and largely unalterable aspect of our human nature. This evolutionary basis supports the claim that social comparison processes underpinning the BFLPE are pan-human and universal (Marsh & Seaton, 2015).

An integrative meta-theory and statistical model

In the present study, we extend the traditional process model of social comparison and integrate into a single theoretical model: extensions of the BFLPE, the internal/external frame-of-reference (I/E) model, and dimensional comparison theory (DCT). Thus, we extend the notion of psychological comparison to include frames of reference based on performances by the same person in respect to different domains (dimensional comparisons) as well as performances in the same domain by other people (social comparisons). This integrated meta-theory yields new predictions that are distinct from the predictions arising from each theory considered separately, resulting in the prediction of five distinct frameof-reference effects (Figure 1).

Statistical models in BFLPE research are increasingly based on sophisticated multilevel models that juxtapose frame-of-reference effects that occur at the level of the individual student (L1), the school (L2), and, in large crosscultural studies, the country (L3; Marsh & Seaton, 2015). Using PISA 2012 data, we test the generalizability of the results across nationally representative samples of 15-year-olds from 68 countries/regions, and we note that this test of generalizability is akin to traditional meta-analysis tests of generalizability across different studies. We begin with a summary of the different frame-of-reference effects and then move to the present investigation in which we test a new model integrating all five models (Figure 1).

The I/E model: Dimensional comparison effects.

I/E model at individual-student level (Figure 1A)

The differentiation among ASCs in specific school subjects is muchgreater thanfor theassociated measures ofachievement. More specifically, math and verbal self-concepts are nearly uncorrelated, whereas math and verbal achievements are substantially correlated (Marsh, 1986, 2007; Marsh et al., 2014). Marsh (1986) developed the I/E model (Figure 1A) to provide a theoretical rationale for why math and verbal self-concepts were so highly differentiated. The I/E model proposed that ASCs in particular school subjects are based on two distinct comparison processes. The first isan external comparisonprocess; self-concepts are based on objective indicators (e.g., test scores, school grades, and achievement feedback) in specific school subjects. The second is an internal (cross-dimensional) comparison process; students compare their own achievements in one particular school subject (i.e. one dimension) with their own achievements in other school subjects (i.e. other dimensions). Thus, the I/E model (Figure 1A) predicts that paths leading from achievement in one domain (e.g., math) to ASCs in the matching domain [e.g., math self-concept (MSC)] are positive, whereas cross-paths leading from achievement in that domain to ASC in a contrasting domain (e.g., verbal) are negative. There now exists extensive support for these predictions based on a large cross-cultural study (Marsh & Hau, 2004) as well as subsequent metaanalyses (Möller et al., 2009; also Huang, 2011).

DCT: An extension of the I/E model (Figure 1B)

Tests of the I/E model are typically based on math and verbal constructs. DCT (Möller & Marsh, 2013) expands the scope of the I/E model, juxtaposing the effects of multiple achievement domains on the corresponding ASC facets. In this extension of DCT, different components of ASC are posited to vary along an a priori continuum (Marsh & Shavelson, 1985) that ranges from MSC at one end to verbal self-concept at the opposite end, with other core academic subjects falling between these two endpoints. Consistent with the basic I/E model, DCT predicts that for school subjects that are ‘far’ from each other (e.g., math and verbal, as in the original I/E model), the effect of academic achievement in one subject on ASC in the other should be negative (Figure 1B). However, if subjects are ‘near’ each other (e.g., math and physics), then the effect of academic achievement in one subject on ASC in the other should be much less negative, or positive, compared with those based on ‘far’ subjects. Subsequent research has found support for these DCT predictions in relation to ‘near’ and ‘far’ subjects (e.g., Guo, Marsh, Morin, Parker, & Kaur, 2015; Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh et al., 2015; Marsh, Kuyper, Seaton, et al., 2014).

DCT (Möller & Marsh, 2013) places the crossdimensional comparison process posited in the I/E model into a broader theoretical foundation in relation to more general psychological models of self-evaluation, person perception, frames of reference, and psychological comparison processes. For example, dimensional comparison effects have been demonstrated in relation to anxiety (Arens, Becker, & Möller, 2017), physical ability (Arens & Preckel, 2018), learning environments (Arens & Möller, 2016), and personality (Helm, Abele, Mueller-Kalthoff, & Möller, 2017).

The BFLPE and its extensions

BFLPE and school-average level (Figure 1C)

Theoretically, the BFLPE posits that students compare their own academic accomplishments with those of their classmates and use this social comparison as a frame-of-reference for the formation of their ASCs (Marsh, 1991; Marsh et al., 2008; Marsh et al., 2014; Marsh, Kuyper, Morin, Parker, & Seaton, 2014; Marsh & Parker, 1984; Marsh & Seaton, 2015; Nagengast & Marsh, 2012; Zell & Alicke, 2009). Students attending schools in which school-average achievement is high are predicted to have lower ASCs than do students of equal ability attending schools in which school-average achievement is low or medium. Consistent with the traditional definition of contextual effects, the negative effect of school-average achievement is over and beyond the positive effects of individual-student achievement on ASC (Marsh & Seaton, 2015). In the most basic BFLPE model (Figure 1C), individual achievement has a positive effect on ASC, but the effect of school-average achievement is negative.

There now exists a large body of research in support of the BFLPE; in particular, studies based on PISA data support its cross-cultural generalizability (Marsh & Seaton, 2015). Because support for generalizability is an important aspect of the present investigation, we present estimates of the BFLPE (Table 1) on the basis of four PISA data collections between 2000 and 2012 (PISA 2000: Marsh & Hau, 2004: 103 558 students from 26 countries; PISA 2003: Seaton, Marsh, & Craven, 2009, 2010: 265 180 students from 41 countries; PISA 2006: Nagengast & Marsh, 2012: 397 500 students from 57 countries; PISA 2012: Marsh, Parker, & Pekrun, 2019: 485, 490 from 68 countries/regions). In all but one of the 191 samples, school-average achievement had a negative effect on ASC and was significantly negative in 181 samples (also refer to recent meta-analysis of BFLPE studies by Fang et al., 2018). The BFLPE tends to increase during the period that students attend the same high school (Marsh, Köller, & Baumert, 2001) and is maintained after students graduated from high school (Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007). There is also clear evidence that ASC formed in high school predicts long-term educational attainment up to at least 8 years after students have graduated from high school, even after controlling the effects of IQ, socio-economic status, standardized achievement tests, and school grades (Marsh & O'Mara, 2008; also Guo, Marsh, et al., 2015; Guo, Parker, Marsh, & Morin, 2015).

We also note that owing to the separate theoretical developments of the I/E model (and DCT) and the BFLPE, there has been some conflation in terminology. We differentiate between the two in Section S1. For present purposes, we use the term ‘external comparison’ from the original I/E model to refer to effects of achievement in one domain to ASC in the same domain (e.g., effect of math achievement on MSC; Figure 1A and B), and we distinguish this from social comparison processes that are at the heart of the BFLPE (the effect of school-average achievement in Figure 1C and D).

The ‘bright student’ hypothesis (Figure 1D)

Extensive research has been performed on the effect of L1 student achievement as a moderator—a potential protective factor—of the negative effects of the BFLPE for students in high-ability schools. The bright student hypothesis (Coleman & Fults, 1985; Huguet et al., 2009; also Marsh & Seaton, 2015; Marsh et al., 2008; Marsh, Kuyper, Morin, et al., 2014) posits that the highest-achieving students should not suffer BFLPEs. The rationale for this proposal is that the brightest students in high-ability schools and classes are still ranked as near or at the top of their group so that they should retain high ASCs. Thus, Coleman and Fults (1985) suggested that the negative effects for the BFLPE in gifted education classes should be limited primarily to less able students in gifted classes and that the brightest students should be relatively immune to BFLPEs. In contrast, the BFLPE theory suggests that the BFLPE should be similar for students of different ability levels (Marsh, 1984; Marsh, Kuyper, Seaton, et al., 2014). As the BFLPE is based on the difference between student perceptions of their own ability and the average ability in their school, if the schoolaverage ability increases, then the relative self-perceptions of all students decrease. The critical test of these predictions (Figure 1D) is the effect of the interaction between individual and school-average achievement on ASC. The bright student hypothesis predicts that this interaction should be substantial and positive (that the brightest students in each school will have higher ASCs than predicted by just their individual achievement), whereas BFLPE theory predicts that this interaction should be close to zero. Consistent with BFLPE theory and empirical research (Marsh, Kuyper, et al.; Marsh & Seaton, 2015), but contradicting the bright student hypothesis, BFLPEs have been shown to be consistent over students of different levels of achievement; crosslevel interactions between L1 individual-student achievement and L2 school-average were consistently small in size across many studies and over many countries and were not even consistent in direction.

Red shirting[1] (relative year-in-school) effect at the individual-student level (Figure 1F)

For a variety of different reasons (e.g., acceleration and starting school at an early age), students might be in classes with older, more academically advanced classmates who constitute a potentially more demanding frame of reference than are classmates of the same age. Marsh (2016; also Marsh et al., 2017; Parker, Marsh, Thoemmes, & Biddle, 2019) found that

[1] ‘red shirting’ came from US college sports and referred to holding The term a college-athlete back a year to develop skills and extend eligibility. By way of analogy, in school settings, the term refers to holding students back so that they are older than their classmates.

the effect of relative year in school (red shirting) on ASC was positive (Figure 1F); the effect on ASC of being young relative to classmates was negative, whilst the effect of being old relative to classmates was positive. Put simply, regardless of how students end up being older than their same-grade classmates, the red shirting effect implies that being older leads to higher ASCs, whereas being younger leads to lower ASCs (for further discussion, refer to Section S2). This finding has been demonstrated cross-culturally using PISA 2003 (Marsh, 2016) and PISA 2012 data (Marsh, Parker, & Pekrun, 2019).

Paradoxical country-level frame-of-reference effects (Figure 1E)

Marsh, Parker, and Pekrun (2019) demonstrated that the same frame-of-reference perspective used to explain contextual effects at the individual student (L1: red shirting) and school-average levels (L2: BFLPE) also applies at the macro-contextual country (L3) level. Thus, students in a given country compare their individual achievements with the levels of achievement in their own country rather than with the levels of achievement of students from different countries. Extending the BFLPE and social comparison theory, they demonstrated that country-average achievement has a negative effect on individual-level ASC. Based on comparisons of self-concept and achievement in the USA with that in Japan, China, Singapore, and other East Asian countries (e.g., Shen & Tam, 2006; also Minkov, 2008; Stevenson, Chen, & Lee, 1993; Sanchez & Dunning, 2018), ASC and achievement were found to be positively correlated at the level of the individual student within each country but negatively correlated at the country level. Thus, US students have higher ASCs than do East Asian students, even though the academic achievements of these US students are lower. Marsh, Abduljabbar, Parker, et al. (2014) extended this research to include Saudi Arabian students, who had lower levels of achievement than US students, but even higher levels of ASC. In the most basic test of this prediction (Figure 1E), the predicted effect of individual achievement on ASC is positive, but the predicted effect of country-average achievement is negative.

In positing this psychological comparison process at the country level, we note that students do not actually have to know how achievement in their country compares with that of students in different countries or to make this comparison process explicit. Rather, the comparison process is indirect as it is based on feedback they receive about their academic performance. In some cases, that will be explicit feedback in relation to national norms on standardized (Marsh, Kuyper, Morin, et al., 2014). Furthermore, the feedback that teachers provide to students about their performances is likely to be in relation to performances by other students in the same school and the same country, not students from different countries. In this sense, the process is consistent with Bronfenbrenner's (1979) ecological model such that macro-contexts can have their influence on individuals through micro-contexts. If, for example, the mean of school-average achievements in one country is higher than in another country this will mean that each child's school context will, on average, also be higher and explains the link between country-average achievement and self-concept (remembering that we condition on individual performance on the same achievement test in each country and thus are comparing equally able children). This is analogous to the BFLPE at the school level, but occurs at the level of the country.

Integration of BFLPE and I/E models: BFLPE-compensatory effect (Figure 1G)

According to the I/E model, students use their own achievement in one area (e.g., verbal) as a frame of reference for the formation of their self-concept in a different area (e.g., math). Thus, the effects of achievement on self-concept in a contrasting domain are opposite to effects of achievement in a matching domain. According to the BFLPE, students use the average accomplishments of their classmates in one area (e.g., math) as a second basis for the formation of their self-concepts in the same domain. Applying the logic of the I/E model to the BFLPE results in the prediction that school-average math achievement has a negative effect on MSC but that the effect of schoolaverage verbal achievement should be positive (compensatory effects model, Figure 1G; Marsh 1984; also Guo, Marsh, Parker, & Dicke, 2018; Marsh, Parker, & Craven, 2015; Parker, Marsh, Lüdtke, & Trautwein, 2013). Strickhouser and Zell (2015) provided support for these predictions using experimentally manipulated feedback to control both frame-of-reference effects in relation to other people and other domains. Results indicated that both cross-dimensional and social comparisons influenced selfevaluations and affective reactions. Further extending this integration of the I/E and BFLPE models to incorporate DCT suggests that the effects of a subject that is near to math (e.g., science) on MSC would fall somewhere between the negative effects of school-average math achievement (BFLPE) and the positive effects of schoolaverage verbal achievement [BFLPE-compensatory effect (BFLPE-CE)].

It is also relevant to explore the psychological processes that might underlie this theoretical prediction. In the BFLPE, children come to think of their abilities in a given domain more poorly if they are in a school where other students do very well in this domain. The I/E model suggests that if children come to doubt their abilities in one domain, such as math, they might come to view their abilities in other fields more positively owing to changes in their internal ranking of their abilities in different domains. Thus, where school average ability leads children to doubt their math ability, this will likely affect the position of math in children's relative internal ranking of their strengths, leading them to evaluate their verbal abilities more positively. Parker et al (2013) provided a demonstration of this process in relation to magnet schools where students are selected based on their performance in a particular academic domain. Once selected, such children are faced with an environment where all children are very good in this particular domain. As such, these children come to evaluate their ability more positively in domains that the magnet school does not target for selection.

Fully integrated social and dimensional comparison models (Figure 1H)

In the fully integrated social and dimensional comparison model (Figure 1H), all the effects are integrated into a single three-level model (L1 = individual student, L2 = school, and L3 = country). In this model, we simultaneously test a priori predictions on the basis of the I/E model (Figure 1A), DCT (Figure 1B), the BFLPE (Figure 1C), the bright student hypothesis (Figure 1D), the paradoxical cross-cultural effect (Figure 1E), the relative year-in-school effect (Figure 1F), and the BFLPE-CE (Figure 1G). Combining all the effects into a single model provides a parsimonious test of the entire set of predictions but also provides strong tests of the distinctiveness of the different effects. Importantly, bringing these disparate models into a single integrated model results in new predictions that are not posited in the models considered separately (e.g., BFLPE-CE) and helps to clarify terminology that has been used inconsistently in the different models.

Cross-cultural generalizability of psychological comparison effects

Interest in cross-cultural approaches to personality peaked in 1948 (Kluckhohn & Murray, 1948) but declined thereafter (refer to systematic review by Chiu, Kim, & Wan, 2008). In the relatively few cross-cultural studies published in major personality journals, the focus has typically been on cross-cultural differences rather than cross-cultural generalizability (Chiu et al., 2008; Chiu, 2012). Marsh, Xu, and Martin (2012; Segall, Lonner, & Berry, 1998; Van de Vijver & Leung, 2000) emphasized that in crosscultural research, there are two main contrasting perspectives. Cultural relativist (idiographic, emic, and qualitative) perspectives accentuate the individual's uniqueness that undermines comparison. In contrast, universalist (nomothetic, etic, neo-positivist, and quantitative) perspectives accentuate cross-cultural similarities. In universalist perspectives, as in the present investigation, the focus is on empirical tests of theoretical predictions, replicability of findings, and cross-cultural generalizability. Hence, one of the goals of cross-cultural research is to test the replicability of existing theories in other cultures, investigate new perspectives in diverse cultural contexts, and propose universal principles, 'in order to generate more nearly universal psychology, one that has pan-human validity' (Segall et al., 1998, p. 1102). Matsumoto (2001) argued that 'cultural differences challenge mainstream theoretical notions about the nature of people and force us to rethink basic theories of personality, perception, cognition, emotion, development, social psychology, and the like in fundamental and profound ways' (p. 9).

More broadly conceived, the universalist approach to cross-cultural research is an example of a 'multiple method' approach to construct validity (Marsh, Martin, & Hau, 2006) that focuses on convergence in relation to the different countries in cross-cultural research. The generalizability of support for a priori predictions based on strong theoretical models, across a diverse set of countries, provides a strong basis of support for the construct validity of theoretical predictions and their interpretations. Here, we adapt the universalist perspective of cross-cultural research to test the generalizability of our a priori predictions across different countries. Following from Van de Vijver and Leung (2000), we note that strong cross-cultural universalist studies require comparable samples and measures from many countries. In this way, differences in the composition of samples and materials are not confounded with cross-cultural differences. PISA data provide a unique opportunity to test the cross-cultural generalizability of theoretical predictions. Cross-cultural comparisons based on databases such as PISA provide strong tests of the generalizability and external validity of theories, measures, models, and a priori predictions.

In summary, many of the traditional limitations and challenges to the universalist approach to cross-cultural research are largely overcome through the use of PISA data, although there are still limitations in relation to cultural relativist (emic) perspectives. It is also relevant to note that this approach to cross-cultural generalizability shares many features of conventional meta-analyses, which have been the traditional approach to testing the robustness and generalizability of support for theoretical predictions.

The Present Investigation

In this article we test predictions in relation to the meta-theoretical integration of the I/E model and dimensional comparison theory (Hypothesis 1) with BFLPE and social comparison theory (Hypothesis 2), forming the big-fish-little-pond-compensatory effect (BFLP-CE) model (Hypothesis 3).

Hypothesis 1: I/E model and DCT at the individual student level (Figures 1A & 1B)

A. The effect of individual student math achievement on MSC will be positive, but the effect of reading achievement on MSC will be negative (I/E model; Figure 1A).
B. The effect of science achievement on MSC will be less negative (closer to zero) than the negative effect of reading achievement on MSC, and less positive (closer to zero) than the positive effect of math achievement on MSC (DCT; Figure 1B).

Hypothesis 2: BFLPE and social comparison theory at the school-average level

A. The effect of individual student (L1) math achievement on MSC will be positive, but the effect of school-average math achievement on MSC will be negative (BFLPE, Figure 1C).
B. There will be little to no support for the "bright student" hypothesis, that the BFLPE will be smaller for more able students and larger for less able students (i.e., the interaction between L1 and L2 achievement will be small or non-significant, and not substantially positive, as the “bright student” hypothesis would predict; Figure 1D).
C. The effect of country-level achievement on MSC will be negative, providing a theoretical explanation of the paradoxical self-concept country effect (Figure 1E).

D. The effect of relative year in school (an individual student's year in school relative to that of same-age classmates) on MSC will be positive (Figure 1F).

Hypothesis 3: Big-fish-little-pond compensatory effect (BFLPE-Compensatory effect), integration of dimensional and social comparison theories at the school-average level (Figure 1G)

A. Following from the logic of the I/E model (Hypothesis 1A), the BFLPE-Compensatory effect predicts that the effect of school-average (L2) reading achievement on MSC will be positive (i.e., opposite in direction to the negative effect of school-average math achievement, the BFLPE).

B. Following from the logic of the dimensional comparison theory extension of the I/E model (Hypothesis 1B), the effect of school-average science achievement on individual MSC will be less positive (closer to zero) than the positive effect of school-average reading achievement on MSC, and less negative (closer to zero) than the negative effect of school-average math achievement on MSC (i.e., the effect of L2 science achievement falls between the positive effect of L2 reading achievement and the negative effect of L2 math achievement).

Hypothesis 4: Universalist perspective to cross-cultural generalizability

Based on a three-level multilevel model (L= students, L2 = schools, L3 = country), we posit that the predictions based on hypotheses 1–3 will generalize across the 68 countries/regions included in the PISA2012 database.

METHOD

Sample and variables

Data used in the present investigation are publicly available through the Organisation for Economic Co-operation and Development (OECD) PISA website, along with a detailed description of the data collection methods, sampling procedures, and measures. For the present purposes, we used all available data for the selected variables. Analyses are described in sufficient detail to allow replication of the results, and selected syntax is included in Section S3. Results are presented as standardized effect sizes along with standard errors that allow computation of confidence intervals and exact p-values.

The set of PISA data collections constitutes one of the largest secondary databases in the public domain and has been the basis of a very large number of studies—including several publications relevant to the present investigation, which are summarized in the literature review (Table 1). For the PISA 2012 data used here (485, 490 fifteen-yearold students, 18 292 schools, 68 countries/regions), the primary focus is on math. PISA data are exemplary in terms of testing and demonstrating the cross-cultural equivalence of measurement across countries (Marsh, Hau, Artelt, Baumert, & Peschar, 2006; also Section S4).

Students completed a survey to assess student and family background variables, and a variety of psychosocial variables, like self-concept, with a focus on math, as well as paper-and-pencil tests of skills and knowledge in reading, math, and science. Data were collected using a complex twostage sampling design and were nationally representative samples, after using the appropriate survey weights. Because the present investigation is based on secondary data analysis of publicly available data, the study is exempt from ethical consent considerations.

In the PISA database, math achievement is represented by five plausible values designed to prevent biased population estimates and to control for measurement error (OECD, 2014). The squared value of math achievement was used to represent the quadratic effect of achievement (Marsh & Hau, 2003). School-average and country-average achievement were based on individual student (L1) measures of achievement aggregated to the school-average (L2) and country-average (L3) levels, whilst the interaction between L1- and L2-math achievement (Mach and L2Mach) was represented as the cross-product of those two variables (L1 × L2-math achievement). In addition to math achievement, the sets of plausible values were also available for reading and science achievement.

MSC, the central variable considered here, was measured with five items (I am just not good at mathematics; I get good grades 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). Based on the item response theory, it was represented as the weighted likelihood estimate provided with the PISA data, as recommended in the PISA Technical Report (OECD, 2014). For these data, MSC was correlated with the three achievements (latent correlations of .26 for math, .17 for science, and .12 for reading), whereas the three achievement scores were highly correlated with each other (correlations greater than .86).

Relative year in school is defined as the difference between the student's year in school and the average year in school for same-age students in that country. The definition follows from the PISA construction of this variable in relation to country-average values and operationalizations of the construct in previous research (e.g., Marsh, 2016). Positive values reflect students being in a higher year in school than the average for 15-year-olds in their country, whilst negative values reflect students being in a lower year in school than the average for 15-year-olds in their country.

Analysis

Multilevel modelling was conducted with the MLwiN statistical package (Rasbash, Steele, Browne, & Prosser, 2004; also Marsh, 2016) to accommodate the three-level hierarchical PISA structure: students (L1) nested within schools (L2) and schools nested within countries (L3). Cases were weighted by the weighting variable provided as part of the PISA database (OECD, 2014). The fixed effects included individual (linear and quadratic) achievement, school-average achievement, country-average achievement, and relative year in school (refer to Sections S3 and S4 for syntax and more details). Random effects included the intercepts at the three levels and also country-level residual variances used to evaluate country-to-country variation in the critical frameof-reference effects. Due in part to the PISA 2012 design, in which achievement is represented as five plausible values, we conducted analyses on five imputed data sets and combined them using Rubin's (1987) rules (refer to further discussion in Section S4).

Standardized metric

We standardized individual student scores (M = 0, SD = 1) across the entire sample to facilitate interpretations in relation to a standardized effect-size metric based on a common metric across all students, schools, and countries. However, none of the multiplicative effects (quadratic achievement, L1 × L2Mach interaction) or aggregated variables (L2-math and L3-math achievement; L2-reading and science achievement) were re-standardized: thus, they were kept in the same metric as individual student variables.

RESULTS

Consistent with the integrative nature of the present investigation, tests of each of the effects are based on the final, fully integrated social and dimensional comparison theoretical model (Figure 1H). In the initial presentations of the results, we focus on parameter estimates for the most comprehensive statistical model representing our theoretical integration (Model 3, Table 2), but we subsequently discuss the juxtaposition between the three different statistical models depicted in Table 2. For purposes of presentation, we will refer to figures that depict separate predictions for each component of the model but base conclusions on results from the fully integrated model. However, in Sections S5 and S6, we also present results for a number of intermediate models leading to the final model.

Hypothesis 1: I/E model and DCT at the individual-student level (Figure 1A and B)

Consistent with predictions based on the I/E model (Table 3), the effect of L1-math achievement is positive (β = .91, SE = 0.034), but the effect of L1-reading achievement is negative (β = .32, SE = 0.023). The effect of science achievement is also negative (β = .18, SE = 0.016). However, in line with the DCT extension of the I/E model, this negative effect of science achievement is less negative (closer to zero) than the negative effect of reading achievement but less positive (closer to zero) than the positive effect of math achievement. Indeed, it is relevant to note that the effects of L1 measures of math, reading, and science achievement in Model 3 (Table 2) are nearly the same as the corresponding effects of a model that had only these three individual student achievements as predictor variables (Section S6, Table S2). This is important, as it demonstrates that the dimensional comparison effects are reasonably independent of the three psychological comparison effects included in Model 3. In summary, there is good support for Hypothesis 1 (i.e. the I/E model and its DCT extension).

Hypothesis 2: BFLPE—social comparison theory at the school-average level Hypothesis 2A: BFLPE (Figure 1C)

We begin with the traditional BFLPE (Figure 1C), in which MSC is predicted by math achievement at the individual student (L1) and school (L2) levels. In keeping with previous BFLPE studies based on PISA data, the quadratic effect of math achievement (L1) was included to account for a small non-linearity in the relation between achievement and selfconcept. Consistent with Hypothesis 2A, the effect of L1-math achievement was positive (as reported under Hypothesis 1), whilst the effect of L2-math achievement (the BFLPE) was negative (β = .50, SE = 0.027; Model 3, Table 2).

Hypothesis 2B: Bright student hypothesis (Figure 1D)

Consistent with BFLPE theory, as predicted in Hypothesis 2B, there was no support for the bright student hypothesis: the effect of the interaction between L1 and L2 achievement was close to zero (β = .05, SE = 0.015). Indeed, although this interaction effect is small in size, it is statistically significant and in the opposite direction to that predicted by the bright student hypothesis. Thus, the BFLPE is marginally more negative, not less negative, for bright students (i.e. MSC declines rather than increases).

Hypothesis 2C: Extension of the BFLPE and social comparison theory—country-level achievement (Figure 1E)

As predicted by Hypothesis 2C, the effect of country-average math achievement (β = .33, SE = 0.052; Model 3, Table 2) was negative. Students formed their MSCs in relation to levels of achievement in the country where they live, rather than achievement levels in other countries, providing a theoretical explanation for the paradoxical cross-country effect.

Hypothesis 2D: Relative year-in-school effect (red shirting, Figure 1F)

Consistent with Hypothesis 2D, students who were young for their year in school (i.e. started young or skipped a grade) had lower MSCs, whereas students who were old for their year in school (i.e. started late or repeated a grade) had higher MSCs (β = .06, SE = 0.006; Model 3, Table 2).

Hypothesis 3: BFLPE-CE (Figure 1G)

Model 3 (Table 2) also includes school-average measures of reading and science achievement to test new predictions resulting from the integration of social comparison and DCTs into an integrated meta-theory of self-concept formation. Consistent with predictions (Hypothesis 3A), the effect of L2-math achievement is negative (β = .50, SE = 0.027; the

BFLPE), whilst there was a small compensatory positive effect of L2-reading achievement (β = .14, SE = 0.024), which is consistent with the BFLPE-CE. Thus, being in a school with high levels of average math achievement has a negative effect on individual-level MSC (BFLPE), but being in a school with high levels of reading achievement has a smaller, compensatory positive effect on MSC (BFLPE-CE).

The effect of L2-science achievement was consistent with Hypothesis 3B, based on the logic of DCT and its integration into the social comparison theory. The effect of schoolaverage science achievement was close to zero but significantly positive (β = .06, SE = 0.017). However, this effect was less positive (closer to zero) than the positive effect of L2-reading achievement (β = .14, SE = 0.024) but less negative (closer to zero) than the negative effect of L2-math achievement (β = .50, SE = 0.027). In summary, there is good support for Hypothesis 3 and the BFLPE-CE.

Hypothesis 4: Universalist perspective on cross-cultural generalizability

Random variance components

We now turn to two aspects of the models in Table 2 that are relevant to tests of a universalist perspective on cross-cultural generalizability (also refer to Section S7). The first of these is the control for background/demographic covariates. The second is the inclusion of random effect estimates, representing country-to-country variation in the fixed effects.

Model 2 (Table 2) includes five background/demographic covariates: gender (female); immigration status (first-generation and second-generation immigrants compared with natives); and two components of socio-economic status (parental education and occupation). The following effects are of interest: girls have lower MSCs; immigrants have higher MSCs (but the effect is less positive for secondgeneration immigrants than for first-generation immigrants); parental education has a small positive effect on MSC (but parental occupation has no significant effect). However, the main focus of the present investigation is how controlling for these variables influences support for a priori predictions. Inspection of the parameter estimates in Models 1 (without covariates) and 2 (with covariates) demonstrates that the inclusion of covariates has almost no effect on the parameter estimates. In summary, support for Hypotheses 1–3 is robust in relation to control for these background/demographic covariates.

In Model 3, key parameter estimates are made random at the country level. Thus far, the focus has been on fixed effects rather than on random effects. However, to the extent that these random effects are large and statistically significant, there is significant country-to-country variation in the results. Because of the large sample size, nearly all the random effects are significant, from a purely statistical perspective. However, if the random effects are small in relation to the corresponding fixed-effect estimate, there is good support for the generalizability of at least the direction of effects in relation to a priori predictions, even if there is significant country-to-country variation in the exact size of the effect.

As a rough rule of thumb, Marsh (2016) suggested that if the standard deviation of country-to-country variation is less than half that of a fixed-effect estimate in support of an a priori prediction, there is good support for the generalizability of the prediction. The rationale for this suggestion is that the direction of the effect will not change even at relatively extreme values (i.e. an individual L1 value that is two standard deviations from the mean of L1 values). For example, consider the results in Model 3 (Table 2) that were used to test Hypothesis 1A (the negative effect of L1-reading achievement on MSC). For this prediction, it is relevant to juxtapose the fixed-effect estimate (β = .32) with the standard deviation of the country-to-country variation (i.e. the square root of the corresponding random variance component, $.03^{1/2}$ = .17). Because the β = .34 is large relative to the SD = 0.17, there is support for the generalizability of at least the direction of the effect across countries and, thus, the a priori prediction.

Of particular relevance is the generalizability of each of the fixed effects representing frame-of-reference effects: L2math achievement (BFLPE); L1-relative year in school (redshirting); L1-reading achievement (dimensional comparison effects); and L2-reading achievement (BFLPE-CE). As noted earlier, these fixed effects (βs = .50; .06; .32; .14, respectively) are all highly significant from a statistical perspective. Compared with country-to-country variation (SDs = $0.027^{1/}$ $_{2\ 1/2\ 1/2\ 1/2}$

= 0.16; .006 = 0.07; .023 = 0.15; .024 = 0.15, respectively), there is reasonable support for the generalizability of the negative effect of L2-math achievement (BFLPE) and the negative effect of L1-reading achievement (DCT) and moderate support for the generalizability of the positive effects of relative year-in-school L2-reading achievement. In summary, an evaluation of the random variance components is largely supportive of the generalizability of these frame-of-reference effects.

Fixed effects for each country: a vote-counting approach An alternative perspective on generalizability is to consider the estimated effects for each of the 68 countries/regions. In Table 3, we summarize the estimated effects for the four frame-of-reference effects considered earlier (also refer to Section S6, Tables S3 and S4 for a summary of all effects). Consistent with observations for country-to-country variation based on the random effects discussed earlier, the negative effects of L2-math achievement (the BFLPE) are nominally significant (more than two SEs below zero) for all 68 countries/regions. The negative effects are almost all significant for both relative year in school (61 of 68 countries) and L1-reading achievement (66 of 68 countries), and none of these 136 effects are significantly positive. However, the results are more mixed for the positive predicted effects of L2-reading achievement: only 42 of 68 predictions are significantly positive, although none is significantly negative.

We also note that part of the reason why more of these L2-reading effects are not significant is that the SEs based on school-level variables are systematically larger than those for the L1-variables (because the number of schools is fewer than the number of students). This is also evident in the SEs for L2-math achievement but is not so critical, as the fixed effects were all substantially larger. Thus, even though the effect sizes of L2-reading achievement were larger than those for L1-year in school, the effects of relative year in school were more consistently significant. In summary, as with the evaluation of the random variance components, an evaluation of the estimates for individual countries is largely supportive of Hypothesis 4, that is, the cross-cultural generalizability of these frame-of-reference effects from a universalist perspective.

DISCUSSION

Given the importance of self-concept (refer to earlier discussion), it is crucial to study processes leading to a positive self-concept. Here, we pursued this aim in relation to an integrated set of frame-of-reference effects based on psychological comparison processes. Juxtaposing individual, group, and even country-level contextual variables, our study contributes to a better understanding of the influences of individual difference variables as well as the impact of social groups and associated contextual effects. The results suggest the relevance of frame-of-reference and group-level contextual effects in personality and social psychology as well as educational psychology, but also to other sub-disciplines of psychology and applied research more generally.

We note that major theoretical models (I/E and BFLPE) have tended to be developed in isolation of each other. Hence, the integration of the five different frame-of-reference effects serves to further the claim that psychological comparison processes are universal phenomena and underline the importance of considering their implications when policy and practice are being formulated (Frank, 2012). BFLPE findings do not support the educational policy in many countries about placement of students in special education settings at both ends of the ability continuum (i.e. negative effects associated with gifted education programs and mainstreaming academically disadvantaged students). Short-term strategies related to schoolstarting age, retention, and promotion are given considerable attention byparents, policymakers,andresearchers. However, red shirting research shows that the effects of all these seemingly disparate strategies can be explained largely in terms of relative year in school and points to benefits associated with school retention when students have not adequately mastered academic materials (Marsh, Pekrun et al., 2017).

Personality research typically has failed to embrace a multidimensional perspective to self-concept and cognitive processes, such as the psychological comparisons that individuals use to form self-perceptions in different domains. Here, we showed why these structural and dynamic aspects are important. Central to this presentation were psychological comparison processes, which are fundamental to all areas of psychology (Festinger, 1954) and which have a universal, evolutionary basis (Frank, 2012). More specifically, we demonstrated five seemingly paradoxical frameof-reference effects in the formation of ASC that were derived from an extended representation of psychological comparison processes.

Only two of these have been considered extensively in previous self-concept research: the negative effects of school-average ability on MSC (BFLPE) and the negative effect of reading achievement on MSC (I/E model). In the present research, we developed a meta-theory, integrating, for the first time, extensions of the BFLPE and social comparison theory, and extensions of the I/E model and DCT, into a single theoretical model based on an extended representation of psychological comparison processes. This integrated meta-theory results in new predictions that are distinct from predictions arising from each model considered separately, producing five distinct frame-of-reference effects.

We also developed multilevel statistical models to test each of the frame-of-reference effects considered separately, as well as a unified statistical model that enables simultaneous testing of each of the five proposed frame-of-reference effects. This provides parsimony in relation to both theory and statistical analysis. However, it also demonstrates how each effect is distinct from the others, results in new predictions based on the integration of different models, and clarifies potential confusion in terminology used in different models. Using PISA 2012 data, we demonstrated the generalizability of support for predictions for all five frame-ofreference effects across nationally representative samples of 15-year-olds from 68 countries/regions. From a universalist perspective of cross-cultural psychology, the results support the cross-cultural generalizability of each of the five frameof-reference effects based on psychological comparison processes.

A meta-theoretical integration of social comparison theory and DCT

The major theoretical contribution of this study is the integration of different models of self-concept formation based on psychological comparison processes that provide theoretical explanations for five frame-of-reference effects. These include extensions of the basic I/E model (Figure 1A and B) and extensions of the basic BFLPE (Figure 1C–F), and their integration to form the BFLPE-CE model (Figure 1G).

I/E and DCT extensions (Hypothesis 1)

The basic I/E model (Figure 1A) proposes a different sort of psychological comparison effect in which the frame of reference is based on performance in one domain relative to performances in other domains. In keeping with the considerable research on this frame-of-reference effect, the effect of math achievement on MSC was substantial and positive, but the effect of verbal achievement on MSC was negative (Hypothesis 1A). It is also consistent with recent extensions of this theoretical model into DCT, which hypothesizes that the effects of science achievement should fall between (i.e. closer to zero) the positive effects of math achievement and the negative effects of verbal achievement (Hypothesis 1B). Further, previous research on DCT, in relation to different science disciplines suggests, for example, that biology falls closer to the verbal end of the self-concept continuum, whereas chemistry andphysics in particular fall closer to the math endof that continuum (e.g., Guo, Marsh, et al., 2015; Jansen et al., 2015; also refer to the Helm et al., 2017, study of perceived subject similarity).

BFLPE extensions (Hypothesis 2)

Three of the frame-of-reference effects are based on extensions of the BFLPE (Hypothesis 2, Figure 1C) and social comparison theory. Emphasizing the theoretical and statistical importance of multilevel modelling, each of these three frame-of-reference effects occurred at a different level of the data: the positive relative year-in-school effect (‘red shirting’) at the L1 level of the individual student (Hypothesis 2C, Figure 1F), the negative effects of school-average achievement (BFLPE at the L2 level of the school, Hypothesis 2A, Figure 1C), and the negative effect of countryaverage achievement (paradoxical cross-cultural effect at L3 level, country; Hypothesis 4). Although support for the BFLPE is widespread, there is little research on either red shirting or country-average achievement, particularly from a frame-of-reference perspective.

Superficially, these three frame-of-reference effects appear to be conceptually distinct, each having discrete implications for policy/practice. However, all three are consistent with straightforward extensions of the BFLPE and social comparison theory. Controlling for individual student achievement, students have lower ASCs when they are in countries where country-average achievement is higher, in schools where the school-average achievement is higher, and in classes where their same-grade classmates are older and more advanced. In each case, students form their ASC relative to their own country, their own school, or their own same-grade classmates. Hence, a theoretically important contribution is the demonstration that all three involve extensions of the same basic BFLPE and social comparison processes that have been the basis of so much past research.

Integrating BFLPE, IE, and DCT (Hypothesis 3)

The final frame-of-reference effect, BFLPE-CE, comes from the integration of the social comparison theory and DCT that results in new predictions (refer to earlier discussion). Thus, whilst the effect of school-average math achievement on MSC (BFLPE) is negative, the effect of school-average verbal achievement (BFLPE-CE) is predicted to be positive (i.e. opposite in direction to that of school-average math achievement; Hypothesis 3A). Furthermore, consistent with the DCT extension of the I/E model, the effect of schoolaverage science achievement is predicted to fall between the negative effect of school-average math achievement and the positive effect of school-average verbal achievement (Hypothesis 3B). The results of the present investigation provide support for both new predictions based on this BFLPE-CE integration of social and DCTs.

A universalist perspective on cross-cultural generalizability and variability (Hypothesis 4)

BFLPE: Cross-cultural generalizability

Seaton et al. (2009) argued that cross-cultural studies like those summarized in Table 1 demonstrate that the BFLPE is one of psychology’s most universal findings, suggesting that it is a pan-human phenomenon. Furthermore, although beyond the scope of the present investigation, in the original model of the BFLPE (Marsh, 1986; also Marsh, 2007; Marsh et al., 2001; Nagengast & Marsh, 2011; Zell & Alicke, 2009), it was proposed that the size of the BFLPE would vary with the sizes of the contextual differences that were posited to underpin it. In relation to cross-cultural comparisons, the size of the contextual differences within each country is a function of the amount of implicit or explicit tracking and variation among schools in terms of school-average achievement. Thus, in relation to differences in the BFLPE across countries, for example, the BFLPE model predicts that there would be no school-level BFLPE in a country if the school-average achievements were the same for all schools within that country. More recently, empirical support for this prediction comes from the finding that the size of the BFLPE in different countries varies substantially with the within-country variation in school-average achievement (Salchegger, 2016; also Marsh, 2007; Marsh et al., 2001; Nagengast & Marsh, 2011; Parker, Marsh, Thoemmes, & Biddle, 2019).

Cross-cultural generalizability of frame-of-reference effects In the universalist perspective in cross-cultural psychology, there is an emphasis on replicability of results, empirical tests, and cross-cultural generalizability of support for theoretical predictions. From a universalist perspective, the remarkable cross-cultural support for all five frame-ofreference effects suggests that the meta-theoretical model upon which they are based constitutes a cross-culturally valid theory of ASC formation. In support of the crosscultural generalizability of the results, it is relevant to juxtapose the large-scale cross-cultural research used here with traditional meta-analyses approaches typically used to test generalizability. Increasingly, both use evolving multilevel analyses that focus on overall ESs as well as residual variance (study-to-study or country-to-country variation). Meta-analysis, despite its many strengths, typically suffers from the heterogeneity of studies in respect to materials, participants, measures, research designs; publication bias; over-reliance on middle-class, Western participants; and over-representation of studies published in English-language journals. Importantly, because meta-analysts rarely have access to individual-level data, appropriate multilevel tests (e.g., effects of L1, L2, and L3-math achievement on L1 MSC) are not possible.

Traditional cross-cultural studies suffer many of the same limitations as meta-analysis. However, the PISA data used here provide a potentially stronger basis for evaluating the cross-cultural generalizability of theoretical predictions than do meta-analyses or traditional cross-cultural comparisons. There are, of course, limitations in the use of PISA data— for example, the age of participants and the cross-sectional nature of the data (refer to subsequent discussion of limitations), as well as the fact that not all countries are included in PISA data collections. Importantly, meta-analyses and large cross-cultural databases are not mutually exclusive such that juxtaposing the results of both approaches within the same study provides stronger tests of generalizability/universality of the findings (Möller et al., 2009). Future substantive-methodological synergies that integrate cutting-edge theoretical development, design, and statistical analyses will provide further insight into these complex issues (refer to related discussion by Hussong, Curran, & Bauer, 2013).

Multilevel perspectives

Multilevel statistical and substantive issues

The multilevel perspective is an important aspect of the present investigation, in that appropriate tests of the statistical significance of these effects require sophisticated multilevel models that control for nesting of persons within groups/institutions, and the nesting of these groups within countries. Substantively, the multilevel perspective is important in that different proposed frame-of-reference effects occurred at the level of individual student (I/E, DCT, and red shirting), school (BFLPE), country (paradoxical self-concept country effect), and even the cross-level interaction of effects at different levels (tests of the bright student hypothesis). The multilevel perspective is also central to tests of the generalizability of effects over countries. We also note, however, that more research is needed to further explore the source of country-to-country variation among these effects. One possibility, for example, is to align countries in relation to spatial proximity, in which countries more geographically close to each other are more similar (e.g., Gebauer et al., 2017), or proximity in relation to other dimensions more relevant to the focus of research such as school policies in relation to tracking. Indeed, as shown by previous research, the most critical variable is the nature of tracking in the school system— operationalized here as variation in school-average achievement (e.g., Marsh & Seaton, 2015). Thus, for example, we already noted that country-to-country variation in the size of the BFLPE is systematically related to the extent of streaming within each country, whereas Huguet et al. (2009; also Marsh, Kuyper, Morin, et al., 2014) showed that when social comparison processes were controlled, BFLPE was substantially truncated.

Relation to Bronfenbrenner's (1979) multilevel ecological model

Here, we consider how contextual processes demonstrated in our article relate to Bronfenbrenner's (1979) ecological model of human development. At the individual level are dimensional comparisons in which children compare their performance in one domain against performance in another. At the micro-system level is the relative year-in-school effects, where parents' choices about when to enrol their child in school and whether to promote or retain them influence children's self-concept. At the meso-system level is the BFLPE, in which school average achievement has a negative influence on self-concept. We place this at the meso-system level as the size of the BFLPE appears to be directly related to the way government policy and sociocultural norms (exo-systems) influence the school context (the micro-system). Simply put, the more a country stratifies its schoolsystem byacademic ability, the larger the BFLPE (Parker, Guo, & Marsh, 2019). Finally, at the macro-system level is the paradoxical country-level effect where it is expected that higher country-average achievement is also negatively related to self-concept.

An evaluation of contextual effects at each level allows us to consider how these levels might interact. For example, considering the meso-system and the individual level, we might expect that school-average achievement in one domain has a positive effect on self-concept in another domain—a BFLPE-CE. Likewise, focusing on context in this way reveals why ‘paradoxical effects’ should be expected. For example, it might seem strange to believe that being retained could have a positive effect on self-concept. However, if it is acknowledged that persistent micro-system influences outweigh the influence of a single discrete event (i.e. the event of being retained in a class), then it is not surprising that retention could have a long-term positive effect on self-concept. Likewise, it seems impossible to believe that children could assess the country-level achievement environment in a way that would impact their self-concepts. However, in considering contextual processes from an ecological perspective, we see that country-level processes do not have a direct effect on self-concept. Rather, their influence comes from macro-system contextual influences on contextual spheres more local to the child.

Strengths, limitations, and directions for further research

We claim to present a meta-theoretical integration of the I/E and BFLPE models to synthesize five single theoretical models. Nevertheless, important limitations in data available in the present investigation (no measures of verbal self-concept and no longitudinal data) mean that there are additional aspects of the formation of ASC that could not be tested with the available data. Here, we discuss implications of some of these limitations and how further extensions could be integrated into the meta-theoretical model posited here. Nevertheless, we emphasize that limitations in the data available to test the model should not be taken as limitations in the theoretical model but merely directions for further research with other data.

Strengths and limitations of PISA data

In the present investigation, the focus is on frame-ofreference effects for 15-year-old students from countries all over the world, in the domain of mathematics. This is a particularly important age, as students from many countries are approaching the end of mandatory education and making critical decisions in relation to further education, training, and work. It is also, of course, easy to argue for the fundamental importance of mathematics (OECD, 2014). However, these foci constitute a double-edged sword in terms of comparability, which facilitates interpretations but also has potential limitations in the generalizability of frame-ofreference effects considered here, to other domains and to different developmental periods. Fortunately, however, there is extensive support for the generalizability of the BFLPE in relation to both domains and age (Fang et al., 2018; Marsh & Seaton, 2015).

Particularly in relation to the I/E model and DCT, it would be important to fully test predictions of the effects of math achievement on self-beliefs in verbal domains, and across a range of school subjects. In this respect, these PISA data are not ideal for fully testing the I/E model. Fortunately, however, there is a growing body of research (refer to earlier discussion of the Möller et al., 2009, meta-analysis) showing that I/E model predictions in relation to the math constructs emphasized here do generalize to those related to verbal constructs, and over different measures, age groups, and countries. Similarly, it is important to explore more fully how the frame-of-reference effects articulated here play out over time in extended longitudinal studies such as the Marsh, Pekrun et al. (2017) study of the development of red shirting over the last 5 years of mandatory schooling in Germany and the transition from primary to secondary schooling.

Particularly where explicit tracking of students into different ability groups occurs within the same school, social comparison effects at the class-average level tend to be even larger than those at school-average level (e.g., Marsh, Kuyper, Morin, et al., 2014)—a phenomenon referred to as the local dominance effect (also Alicke, Zell, & Bloom, 2010; Zell & Alicke, 2009). However, this distinction between social comparisons at the class and school levels has not been incorporated into large cross-cultural studies, because—like the present investigation—the sampling frame for such studies is typically the school rather than intact classes within schools. We note that because of this limitation, the sizes of contextual effects in cross-cultural studies like our study are likely to underestimate the true contextual effects (refer to discussion by Marsh, Kuyper, Morin, et al., 2014).

The cross-sectional nature of PISA data means that we were not able to include temporal comparison processes in our integrated model. Following Marsh (1990), a large number of self-concept studies have tested a reciprocal effects model (REM) showing that ASC and achievement are reciprocally related over time (Marsh & Martin, 2011; also Huang, 2011; Valentine, DuBois, & Cooper, 2004). In traditional personality research, the classical essentialist view is that personality should be relatively immune to the influences of situation, life events, contexts, and environmental effects McCrae (2000; 2004), but such an extreme position is unlikely to be supported (Marsh, 2008; also Fleeson, 2001; Judge, Simon, Hurst, & Kelley, 2014). Thus, McAdams and Pals (2006) argued that personality traits and characteristic adaptations are likely to be reciprocally related. More specifically, Marsh (2008) speculated that over time self-concept and traditional personality factors such as the Big Five are likely to be reciprocally related over time, although the paths from personality traits to self-concept factors are likely to be stronger than vice versa. Furthermore, the causal effects of decontextualized personality traits on contextualized outcomes are likely to be substantially mediated, or even moderated, by domainspecific self-concept factors.

The REM has also been integrated with the I/E model (REM-I/E; Marsh, 1989; Marsh & Köller, 2004; Möller & Marsh, 2013) in studies that did not incorporate social comparison processes inherent in the BFLPE (also refer to the related 2I/E model proposed by Wolff, Helm, Zimmermann, Nagy, & Möller, 2018). Recently, Marsh, Parker, and Pekrun (2019) posited a ‘cube’ model that included all three (temporal, dimensional, and social) comparison processes. Based on longitudinal data from a single country, they provided support for the reciprocal effects of ASC and achievement, the I/E model, and the BFLPE within a single integrated model, but they were not able to test the cross-cultural generalizability, the paradoxical cross-cultural effect, or the BFLPE-CE demonstrated here.

Assumptions of causality and underlying processes

A potential limitation of the present investigation is that it is based on a single wave of cross-sectional data; thus, causality cannot be inferred. Psychological, cross-cultural, and educational studies routinely must rely on non-experimental data, as it would be problematic and also unethical to randomly assign students to different schools, let alone different countries. Although causal relations are appropriately hypothesized, claims of support for these hypotheses need to be interrogated in relation to the construct validity of their interpretations (Marsh, 2007). This requires convergence grounded on multiple investigations, study designs, measures, and time points; and support for generalizability across diverse settings and countries. However, particularly for the BFLPE and I/E models, various combinations of longitudinal, quasi-experimental, matching and true experimental designs, and even introspective diary studies have consistently provided support for a priori predictions that provide a stronger basis for evaluating the causality implied in the theoretical predictions. In Section S8, we outline some of these alternative approaches. In particular, for both the BFLPE and I/E models, there is a growing number of studies that support the predicted effects based on experimentally manipulated feedback prompting social comparison (BFLPE studies; e.g., Alicke et al., 2010; Zell & Alicke, 2009) or comparison across multiple dimensions (I/E studies; e.g., Möller & Köller, 2001; Pohlmann & Möller, 2009; also Strickhouser & Zell, 2015).

Implications for personality research

The results of this study have important implications for future personality research, in terms of both structure and processes. These implications involve the better integration of a multidimensional self into personality research, as well as the nature of research questions, theoretical models, and statistical approaches that are also relevant to personality research.

Structurally, personality research has typically adapted a unidimensional approach to self-concept that focuses on self-esteem, the global component of self-concept. However, like personality, self-concept is a highly multidimensional construct and domain specific so that much variance in particular components of self-concept cannot be explained in terms of a single global component. As noted by Swann et al. (2007) and others, this focus on self-esteem in personality research violates the specificity-matching principle, as illustrated by the Marsh, Trautwein, et al. (2006) study relating personality, self-concept, and academic outcomes. The implication for personality researchers is that a multidimensional approach that focuses onspecific domains of self-concept is more useful than a unidimensional perspective. Indeed, a viable compromise might be to consider a multidimensional, hierarchical model of self-concept that juxtaposes global and domain specific components of self-concept (Marsh, Trautwein, et al., 2006). Moreover, this is likely to be true not only for selfconcept but for other personality constructs as well. For example, trait emotions such as anxiety are largely domain specific (Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007; Marsh, 1988; Pekrun & Linnenbrink-Garcia, 2014), but this is typically ignored in classical personality research. As such, personality research on other constructs of personality could follow the model of self-concept research in testing possible multidimensionality.

Particularly within ASC research, there is a major focus on domain specificity. Thus, for example, ASCs in relation to different school subjects are highly differentiated and stable over time, and highly predictive of domain-specific outcomes. An analogous question in personality research would be the manifestation of a particular personality trait as domain specific. Thus, for example, does neuroticism for a particular individual generalize over social, work, academic, sport, and family domains (also Goetz et al., 2007)? Does conscientiousness generalize, or is it specific to work, academic, or sport settings, or even more specific to select domains of work, school, or sports?

From a process perspective, the focus on frame-ofreference and psychological comparison processes has important implications for personality process research. The effects of stable personality traits on distal outcomes of importance are likely to be mediated by self-perceptions and the dynamic processes that individuals use to frame self-perceptions. Hence, McAdams (1996) notes the historical problem personality research has in integrating decontextualized Big Five factors and contextualized outcomes. Indeed, support for an evolutionary basis for the universality of social comparison processes (Frank, 2012) suggests that they are fundamental in understanding the whole person. More generally, there is a need for the self and self-processes to be better integrated into personality research, and for the whole person, which has been separated from social contexts in personality research, to be better integrated with the self, self-perceptions, and self-processes.

Relatedly, there is also another interesting distinction between the interpretation of self-perceptions in self-concept and personality research. Self-concept is inherently based on self-perceptions. Hence, so long as participants truthfully report their self-perceptions, these self-reports are unbiased measures of self-concept even if they are unrealistic in relation to external criteria (i.e. objective accomplishments, inferred self-concept ratings by others; for further discussion, refer to Marsh, 2007). In this respect, self-concept responses might be seen as biased predictors of these external criteria, but unbiased measures of self-concept, whilst the external criteria are potentially biased predictors of self-concept (Marsh, 2007). In contrast, personality researchers typically see self-reports as a potentially biased approach to assess true, underlying personality traits. From this perspective, the influences of psychological comparison processes can be viewed as potential source of bias in relation to true personality traits and, perhaps, less interesting. A compromise might be to treat self-perceptions of personality traits as true self-perceptions that are important in their own right, even if not completely accurate representations of underlying personality traits. In this way, personality researchers would be better able to assess psychological comparison processes that lead self-perceived to differ from ʻactualʼ personality. This would be useful because psychological comparison processes and frames-of-reference are likely to mediate the effects of underlying personality on self-perceptions, and self-perceptions are likely to mediate the effects of personality on many choice behaviours and outcomes of interest to personality researchers.

In self-concept research, there is an increasing emphasis on cross-cultural approaches to test the generalizability of theoretical predictions from a universalist perspective. Although there is some cross-cultural personality research on the structure of personality (Chiu et al., 2008; Chiu, 2012), self-concept research provides a possible exemplar for testing the cross-cultural generalizability processes and variables that influence personality traits.

Various meta-theoretical models of personality have been developed in the literature (Cloninger, Svrakic, & Przybeck, 1993; Matthews, 2018; McAdams & Pals, 2006; McCrae & Costa, 1996; Murtha, Kanfer, & Ackerman, 1996). Whilst they all differ in various nuances, most distinguish between traits (supposedly stable general tendencies) from characteristic adaptations (expressions of underlying personality traits in context). Domain-specific self-concept is a quintessential characteristic adaptation (Parker et al., 2008). Although some personality research has considered the link between personality traits and characteristic adaptations like self-concept (e.g., Vasalampi et al., 2014), less research has considered the processes by which context influences characteristic adaptations.

Part of the difficulty in conducting such research in relation to personality is determining embedded environments that have sufficient regularity to consider the effect of context at multiple levels. This is why education is such a valuable setting for personality research. In educational settings, these processes are relatively easy to study because there are clear assessment criteria and clear identification of frames of reference (i.e. big differences among schools in relation to school-average achievement), and because the context is highly salient, stable, and not artificial (as in some psychology experiments in which short-term contextual effects are mimicked with false feedback). Almost all children in developed countries attend schools, and these environments are typically regulated in such a way that researchers can identify meaningful variance in contexts at the within-school and between-school levels as well as differences in environments across jurisdictions and countries. Similarly, Diener and Fujitaʼs (1997) review of social comparison studies noted that school settings provided the clearest tests of psychological comparison processes because the frame-of-reference effects based on classmates within the same school were the most clearly defined and ecologically valid, compared with other research settings.

In summary, there are many recent advances in self-concept theory and research that have relevance to personality research. As emphasized here, psychological comparison processes in the formation of self-perceptions drive important choice behaviours and are likely to mediate the effects of personality traits on critical outcomes. Furthermore, self-perceptions of personality traits themselves are likely to be influenced by contextual effects like those considered in the BFLPE and I/E models, and this has important ramifications for personality assessment and applied personality research. To what extent are personality selfperceptions in one area influenced by levels in other areas (refer to Helm et al., 2017 study of agency and communion)? To what extent are self-perceptions of traditional personality traits influenced by the levels of these traits in other individuals in one's local context (do individuals see themselves as more or less neurotic when in a neurotic context)? In order to explore these multilevel contextual effects, self-concept research has developed increasingly sophisticated statistical models that are aligned with substantively important issues, micro-level theoretical models of self-concept formation, and macro-level theoretical models that are also relevant to personality researchers. Clearly, the exploration of all these possibilities is beyond the scope of any one investigation, but the present investigation provides some important answers in relation to self-concept and potentially fruitful directions for personality research to explore.

REFERENCES

Albert, S. (1977). Temporal comparison theory. Psychological Review, 84, 485–503. https://doi.org/10.1037/0033-295X.84.6.485.

Alicke, M. D., Zell, E., & Bloom, D. (2010). Mere categorization and the frog pond effect. Psychological Science, 2, 174–177. https://doi.org/10.1177/0956797609357718.

Allport, G. W. (1937). Personality: A psychological interpretation. Oxford, England: Holt.

Allport, G. W. (1968). Is the concept of self necessary? In C. Gordon, & K. Gergen (Eds.), The self in social interaction (pp. 25–32). New York, NY: Wiley.

Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety: Within- and cross-domain relations with achievement and the mediating role of academic self-concept. Contemporary Educational Psychology, 51, 240–252. https://doi.org/10.1016/j.cedpsych. 2017.08.005.

Arens, A. K., & Möller, J. (2016). Dimensional comparisons in students' perceptions of the learning environment. Learning and Instruction, 42, 22–30. https://doi.org/10.1016/j. learninstruc.2015.11.001.

Arens, A. K., & Preckel, F. (2018). Testing the internal/external frame-of-reference model with elementary school children: Extension to physical ability and intrinsic value. Contemporary Educational Psychology, 54, 199–211. https://doi.org/10.1016/j. cedpsych.2018.06.003.

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

Bandura, A. (2006). Toward a psychology of human agency. Perspectives on Psychological Science, 1, 164–180. https://doi.org/ 10.1111/j.1745-6916.2006.00011.x.

Baumeister, R. F. (Ed) (1999). The self in social psychology. Philadelphia, PA: Psychology Press.

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., ... Jayawickreme, E. (2017). Integrating personality structure, personality process, and personality development. European Journal of Personality, 31, 503–528. https://doi.org/10.1002/per.2115.

Bronfenbrenner, U. (1979). Ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press.

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. https://doi.org/10.1016/j.lindif.2012.07.021.

Chiu, C.-Y., Kim, Y.-H., & Wan, W. W. N. (2008). Personality: Cross-cultural perspectives. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), The SAGE handbook of personality theory and assessment, Vol 1: Personality theories and models (pp. 124–144). Thousand Oaks, CA, US: Sage Publications, Inc.

Chiu, M.-S. (2012). The internal/external frame-of-reference model, big-fish-little-pond effect, and combined model for mathematics and science. Journal of Educational Psychology, 104, 87–107. https://doi.org/10.1037/a0025734.

Clark, A. E. (2018). Four decades of the economics of happiness: Where next? Review of Income and Wealth. Online Series 64, Number 2, June 2018, 64, 245–269. https://doi.org/10.1111/ roiw.12369.

Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A systemic review of interventions for co-occurring substance use disorder and borderline personality disorder. Archives of General Psychiatry, 50, 975–990. https://doi.org/10.1001/archpsyc. 1993.01820240059008.

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, 7–11. https://doi.org/10.1177/074193258500600103.

Costa, P. T. Jr., & McCrae, R. R. (1992). Four ways five factors are basic. Personality and Individual Differences, 13, 653–665. https://doi.org/10.1016/0191-8869(92)90236-I.

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. 329–357). Mahwah, NJ: Erlbaum.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. Educational Psychologist, 44, 78–89. https://doi.org/10.1080/ 00461520902832368.

Elliot, A. J., & Dweck, C. S. (2005). Competence and motivation: Competence as the core of achievement motivation. In A. J. Elliot, & C. S. Dweck (Eds.), Handbook of competence and motivation (pp. 3–12). New York: Guilford Publications.

Epstein, S. (1979). Stability of behavior: On predicting most of the people much of the time. Journal of Personality and Social Psychology, 37, 1097–1126. https://doi.org/10.1037/00223514.37.7.1097.

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

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

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. Journal of Personality and Social Psychology, 80, 1011–1027. https:// doi.org/10.1037/0022-3514.80.6.1011.

Frank, R. H. (2012). The Darwin economy: Liberty, competition, and the common good. Princeton, NJ: Princeton University Press.

Gebauer, J. E., Sedikides, C., Schönbrodt, F. D., Bleidorn, W., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2017). The religiosity as social value hypothesis: Replication and extension. Journal of Personality and Social Psychology, 113, 18–39. https://doi.org/10.1037/pspp0000104.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students′ academic emotions. Journal of Educational Psychology, 99, 715–733. https://doi.org/10.1037/0022-0663.99.4.715.

Graziano, W. G., Jensen-Campbell, L. A., & Finch, J. F. (1997). The self as a mediator between personality and adjustment. Journal of Personality & Social Psychology, 73, 392–404. https://doi. org/10.1037/0022-3514.73.2.392.

Greenwald, A. G. (1988). A social-cognitive account of the self′s development. In D. K. Lapsley, & F. C. Power (Eds.), Self, ego and identity: Interpretative approaches (pp. 30–42). New York: Springer-Verlag.

Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2015). Directionality of the associations of high school expectancy-value, aspirations, and attainment: A longitudinal study. American Educational Research Journal, 52, 371–402. https://doi.org/10.3102/0002831214565786.

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

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. Developmental Psychology, 51, 1163–1176. https://doi.org/ 10.1037/a0039440.

Hair, E. C., & Graziano, W. G. (2003). Self-esteem, personality and achievement in high school: A prospective longitudinal study in Texas. Journal of Personality, 71, 971–994. https://doi.org/ 10.1111/1467-6494.7106004.

Harter, S. (2012). The construction of the self: Developmental and sociocultural foundations. Guilford Press.

Helm, F., Abele, A. E., Mueller-Kalthoff, H., & Möller, J. (2017). Applying dimensional comparison theory to the fundamental dimensions of social judgment–agency and communion. Learning and Individual Differences, 54, 116–125. https://doi.org/ 10.1016/j.lindif.2017.01.011.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. Journal of School Psychology, 49, 505–528. https://doi.org/10.1016/j.jsp.2011. 07.001.

Huguet, P., Dumas, F., Marsh, H. W., Wheeler, L., Seaton, M., Nezlek, J., Suls, J., et al. (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, 671–710. https://doi.org/10.1037/a0015558.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative Data Analysis In Clinical Psychology Research. Annual Review of Clinical Psychology, 9, 61–89. https://doi.org/10.1146/ annurev-clinpsy-050212-185522.

James, W. (1890/1963). The principles of psychology 2. New York: Holt.

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. Journal of Educational Psychology, 107, 1086–1101. https://doi.org/10.1037/ edu0000021.

Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. Journal of Applied Psychology, 99, 199. https://doi.org/10.1037/a0034485.

Kluckhohn, C., & Murray, H. A. (1948) (eds)). Personality in nature, culture, and society. New York: Knopf.

Leary, M. R., & Tangney, J. P. (Eds.)(2003). Handbook of self and identity. New York, NY: Guilford.

Lewin, K. (1935). A dynamic theory of personality. New York: McGraw-Hill.

Marsh, H. W. (1984). Self-concept: The application of a frame-ofreference model to explain paradoxical results. Australian Journal of Education, 28, 165–181. https://doi.org/10.1177/ 000494418402800207.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame-of-reference model. American Educational Research Journal, 23, 129–149. https://doi.org/10.3102/ 00028312023001129.

Marsh, H. W. (1988). The content specificity of math and English anxieties: The high school and beyond study. The International Journal of Anxiety Research, 1, 137–149. https://doi.org/ 10.1080/10615808808248226.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early-adulthood. Journal of Educational Psychology, 81, 417–430. https://doi.org/10.1037/00220663.81.3.417.

Marsh, H. W. (1990). The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. Journal of Educational Psychology, 82, 646–656. https://doi.org/10.1037//0022-0663.82.4.646.

Marsh, H. W. (1991). The failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. American Educational Research Journal, 28, 445–480. https://doi.org/10.3102/00028312028002445.

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

Marsh, H. W. (2008). A multidimensional, hierarchical model of self-concept: An important facet of personality. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), The SAGE handbook of personality theory and assessment, Vol 1: Personality theories and models (pp. 447–469). Thousand Oaks, CA, US: Sage Publications, Inc.

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

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-littlepond effect in mathematics: A cross-cultural comparison of US and Saudi Arabian TIMSS responses. Journal of Cross-Cultural Psychology, 45, 777–804. https://doi.org/10.1177/002202211 3519858.

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

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal-external frame-ofreference predictions across 26 countries. Journal of Educational Psychology, 96, 56–67. https://doi.org/10.1037/0022-

0663.96.1.56.

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. American Psychologist, 58, 364–376. https://doi.org/10.1037/0003066X.58.5.364.

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

Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of east and west German school systems after the fall of the Berlin Wall. Contemporary Educational Psychology, 29, 264–282. https://doi.org/10.1016/S0361-476X(03) 00034-1.

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. https://doi.org/10.3102/00028312038002321.

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. https://doi.org/10.1016/j.learninstruc. 2014.04.0020959-4752/.

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: An extension of the internal/external frameof-reference effect on academic self-concept formation. Contemporary Educational Psychology, 39, 326–341. https://doi.org/ 10.1016/j.cedpsych.2014.08.003.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. Learning and Instruction, 35, 16–32. https://doi.org/10.1016/j. learninstruc.2014.08.005.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. British Journal of Educational Psychology, 81, 59–77. https://doi.org/ 10.1348/000709910X503501.

Marsh, H. W., Martin, A. J., & Hau, K.-T. (2006). A multiple method perspective on self-concept research in educational psychology: A construct validity approach. In M. Eid, & E. Diener (Eds.), Handbook of multimethod measurement in psychology (pp. 441–456). Washington, DC, American Psychological Association: Washington DC.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. Developmental Psychology, 49, 1194–1218. https://doi.org/ 10.1037/a0026913.

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

Marsh, H. W., & Parker, J. W. (1984). Determinants of student selfconcept: 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. https://doi.org/10.1037/ 0022-3514.47.1.213.

Marsh, H. W., Parker, P., & Craven, R. G. (2015). Dimensional comparisons theory: An extension of the internal/external frame-of-reference model. In F. Guay, H. W. Marsh, R. G. Craven, & D. McInerney (Eds.), Self-concept, motivation, and identity: Underpinning success with research and practice (pp. 115–151), International Advances in Self Research (Vol 5). Charlotte, NC: Information Age Publishing.

Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. European Psychologist, 24, 231–242. https://doi.org/ 10.1027/1016-9040/a000332.

Marsh, H. W., Pekrun, R., Murayama, K., Arens, K. A., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over six years. Developmental Psychology, 54, 263–280. https://doi.org/10.1037/dev0000393.

Marsh, H. W., Pekrun, R., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. Journal of Educational Psychology, 109, 425–438. https://doi.org/ 10.1037/edu0000144.

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

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O’Mara, A. J., & Craven, R. G. (2008). The big-fish-littlepond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review, 20, 319–350. https://doi.org/10.1007/s10648008-9075-6.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, 20, 107–125. https://doi.org/10.1207/s15326985ep2003_1.

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

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. Journal of Personality, 74, 403–456. https://doi.org/10.1111/j.1467-6494.2005.00380.x.

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), APA educational psychology handbook, Vol. 1: Theories, constructs, and critical issues (pp. 427–458).

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: Relations to academic self-concept and achievement. American Educational Research Journal, 34, 691–720. https://doi.org/ 10.3102/00028312034004691.

Matsumoto, D. (2001). Cross-cultural psychology in the 21st century. In J. S. Halonen & S. F. Davis (Eds.), The many faces of psychological research in the 21st century. Washington DC: Society of the Teaching of Psychology, Division 2 of the American Psychological Association. Retrieved from http://teachpsych. lemoyne.edu/teachpsych/faces/script/ch05.htm

Matthews, J. G. (2018). Cognitive-adaptive trait theory: A shift in perspective on personality. Journal of Personality, 86, 69–82. https://doi.org/10.1111/jopy.12319.

McAdams, D. P. (1996). Personality, modernity, and the storied self: A contemporary framework for studying persons. Psychological Inquiry, 7, 295–321. https://doi.org/10.1207/ s15327965pli0704_1.

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. American Psychologist, 61, 204–217. https://doi.org/10.1037/0003066X.61.3.204.

McCrae, R. R. (2000). Trait psychology and the revival of personality and culture studies. American Behavioral Scientist, 44, 10–31. https://doi.org/10.1177/00027640021956062.

McCrae, R. R. (2004). Human nature and culture: A trait perspective. Journal of Research in Personality, 38, 3–14. https://doi. org/10.1016/j.jrp.2003.09.009.

McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), Five-factor model of personality: Theoretical perspectives (pp. 51–87). New York: Guilford Press.

Minkov, M. (2008). Self-enhancement and self-stability predict school achievement at the national level. Cross-Cultural Research, 42, 172–196. https://doi.org/10.1177/106939710 7312956.

Mischel, W., & Morf, C. C. (2003). The self as a psycho-social dynamic processing system: A meta-perspective on a century of the self in psychology. In M. R. Leary, & J. P. Tangney (Eds.), Handbook of self and identity (pp. 15–43). New York: Guilford Press.

Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frames of reference model. Journal of Educational Psychology, 93, 826–835. https://doi.org/10.1037/0022-0663.93.4.826.

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

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame-ofreference model of academic achievement and academic selfconcept. Review of Educational Research, 79, 1129–1167. https://doi.org/10.3102/0034654309337522.

Murtha, T. C., Kanfer, R., & Ackerman, P. L. (1996). Toward an interactionist taxonomy of personality and situations: An integrative situational—Dispositional representation of personality traits. Journal of Personality and Social Psychology, 71, 193–207. https://doi.org/10.1037/0022-3514.71.1.193.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries, and the world: The big-fish-little-pond-effect for PISA 2006. Educational Psychology, 31, 629–656. https://doi.org/10.1080/01443410.2011.586416.

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

OECD (Organisation for Economic Cooperation and Development). (2014). PISA 2012 technical report. Paris: Author.

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. Annual Review of Psychology, 57, 401–421. https://doi.org/10.1146/annurev.psych.57.102904. 190127.

Parker, D., Guo, J., & Marsh, H. A. (2019). A macro context theory of academic self-concept: Ability stratification and the big-fishlittle-pond effect. PsyArXiv.

Parker, P. D., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2013). Differential school contextual effects for math and English: Integrating the big-fish-little-pond effect and the internal/external frame-of-reference. Learning and Instruction, 23, 78–89. https://doi.org/10.1016/j.learninstruc.2012.07.001.

Parker, P. D., Marsh, H. W., Thoemmes, F., & Biddle, N. (2019). The negative year in school effect: Extending scope and strengthening causal claims. Journal of Educational Psychology. Advance online publication, 111, 118. https://doi.org/10.1037/ edu0000270.

Parker, P. D., Martin, A. J., & Marsh, H. W. (2008). Factors predicting life satisfaction: A process model of personality, multidimensional self-concept, and life satisfaction. Journal of Psychologists and Counsellors in Schools, 18, 15–29. https://doi.org/10.1375/ajgc.18.1.15.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. Child Development. Advance online publication, 88, 1653–1670. https://doi.org/10.1111/cdev.12704.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In International handbook of emotions in education. New York: Routledge.

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. Journal of Personality and Social Psychology, 117, 166–185. https://doi.org/10.1037/ pspp0000230.

Pohlmann, B., & Möller, J. (2009). On the benefits of dimensional comparisons. Journal of Educational Psychology, 101, 248–258. https://doi.org/10.1037/a0013151.

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

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.

Salchegger, S. (2016). Selective school systems and academic selfconcept: How explicit and implicit school-level tracking relate to the big-fish-little-pond effect across cultures. Journal of Educational Psychology, 108, 405–423. https://doi.org/10.1037/ edu0000063.

Sanchez, C., & Dunning, D. (2018). Cultural patterns explain the worldwide perception/performance paradox in student selfassessments of math and science skill. Social Psychological and Personality Science. https://doi.org/10.1177/1948550618801003.

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-littlepond effect across 41 culturally diverse countries. Journal of Educational Psychology, 101, 403–419. https://doi.org/10.1037/ a0013838.

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-littlepond-effect: Generalizability and moderation—Two sides of the same coin. American Educational Research Journal, 47, 390–434. https://doi.org/10.3102/0002831209350493.

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. https:// doi.org/10.1037/0003-066X.53.10.1101.

Shafer, A. B. (2000). Relation of the big five to biodata and aspects of the self. Personality & Individual Differences, 28, 1017–1035. https://doi.org/10.1016/S0191-8869(99)00126-9.

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

Shen, C., & Tam, H. P. (2006). The paradoxical relationship between students' achievement and their self-perceptions: A cross-national analysis based on three waves of TIMSS data. The Second IEA International Research Conference: Proceedings of the IRC-2006, 1, 43–60.

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. Educational Psychologist, 37, 233–244. https://doi.org/10.1207/S15326985EP3704_3. Stevenson, H. W., Chen, C., & Lee, S. (1993). Motivation and achievement of gifted children in East Asia and the United States. Journal for the Education of the Gifted, 16, 223–250. https://doi.org/10.1177/016235329301600302.

Strickhouser, J. E., & Zell, E. (2015). Self-evaluative effects of dimensional and social comparison. Journal of Experimental Social Psychology, 59, 60–66. https://doi.org/10.1016/j.jesp. 2015.03.001.

Swann, W. B. Jr., Chang-Schneider, C., & McClarty, K. L. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. American Psychologist, 62, 84–94. https://doi. org/10.1037/0003-066X.62.2.84.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. Educational Psychologist, 39, 111–133. https://doi.org/ 10.1207/s15326985ep3902_3.

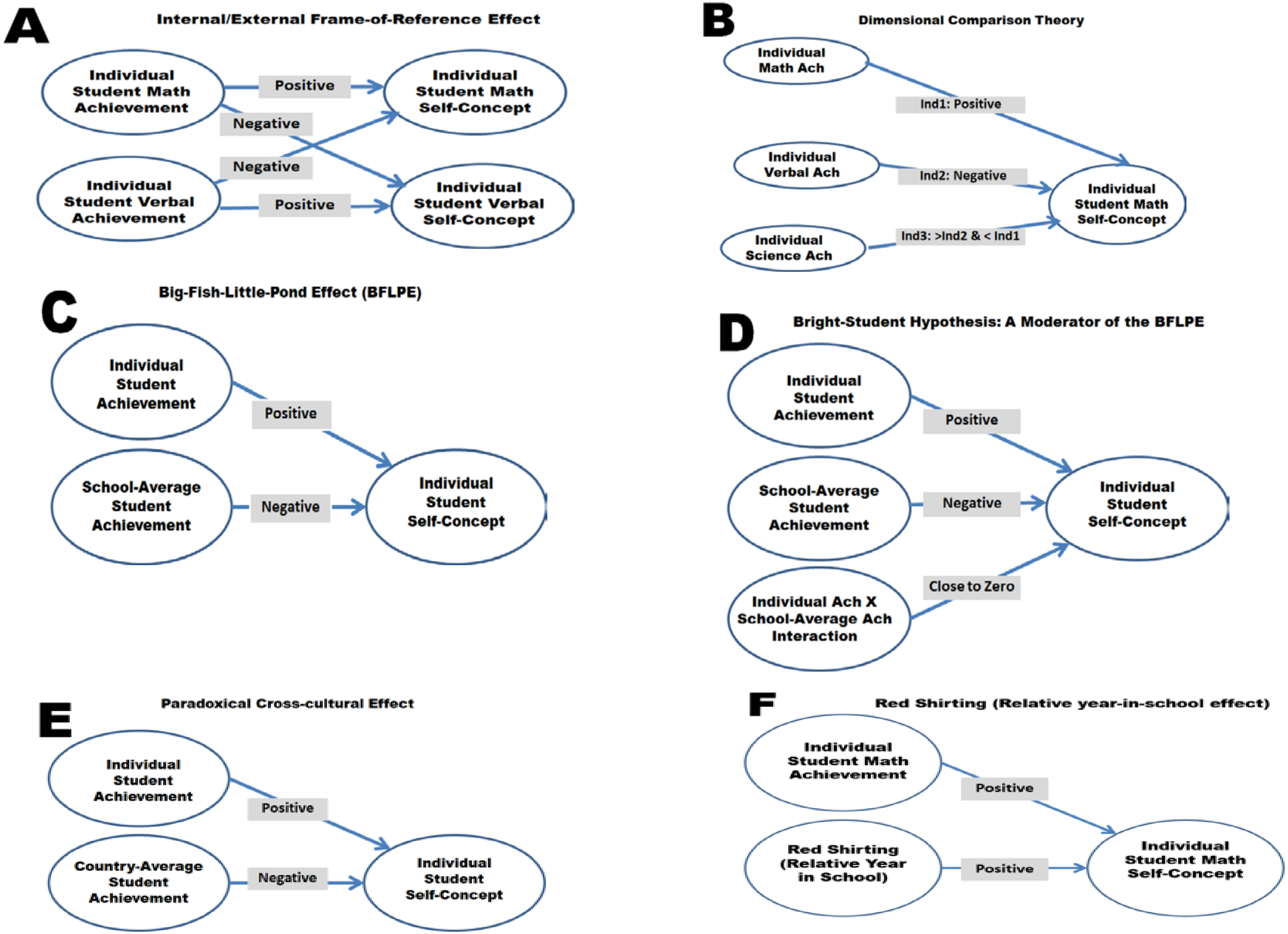
Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. Journal of Cross-Cultural Psychology, 31, 33–51. https://doi.org/10.1177/0022022100 031001004.

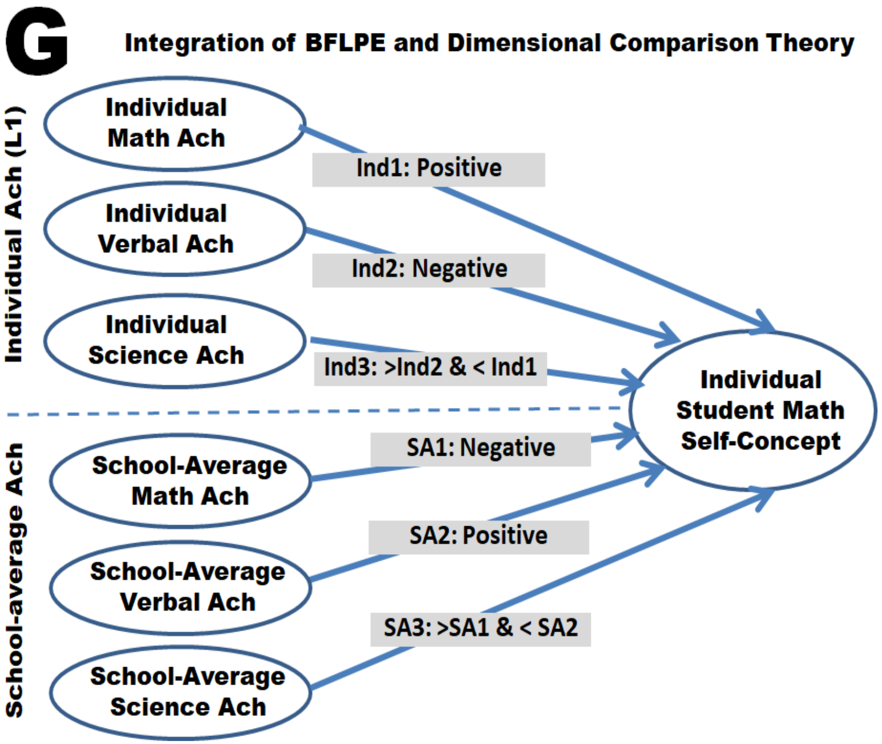
Vasalampi, K., Parker, P., Tolvanen, A., Lüdtke, O., Salmela-Aro, K., & Trautwein, U. (2014). Integration of personality constructs: The role of traits and motivation in the willingness to exert effort in academic and social life domains. Journal of Research in Personality, 48, 98–106. https://doi.org/10.1016/j.jrp.2013.11.004.

Waller, N. G., & Ben-Porath, Y. S. (1987). Is it time for clinical psychology to embrace the five-factor model of personality? American Psychologist, 42, 887–889. https://doi.org/10.1037/ 0003-066X.42.9.887.

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. Journal of Educational Psychology, 110, 1005–1025. https://doi.org/10.1037/edu0000248.

Zell, E., & Alicke, M. D. (2009). Contextual neglect, selfevaluation, and the frog-pond effect. Journal of Personality and Social Psychology, 97, 467–482. https://doi.org/10.1037/ a0015453.


A
Internal/External Frame-of-Reference Effect
Individual Student Math Achievement
Positive
Individual Student Math Self-Concept
Negative
Negative
Individual Student Verbal Achievement
Positive
Individual Student Verbal Self-Concept
B
Dimensional Comparison Theory
Individual Math Ach
Ind1: Positive
Individual Verbal Ach
Ind2: Negative
Individual Student Math Self-Concept
Individual Science Ach
Ind3: >Ind2 & < Ind1
C
Big-Fish-Little-Pond Effect (BFLPE)
Individual Student Achievement
Positive
School-Average Student Achievement
Negative
Individual Student Self-Concept
D
Bright-Student Hypothesis: A Moderator of the BFLPE
Individual Student Achievement
Positive
School-Average Student Achievement
Negative
Individual Student Self-Concept
Individual Ach X School-Average Ach Interaction
Close to Zero
E
Paradoxical Cross-cultural Effect
Individual Student Achievement
Positive
Country-Average Student Achievement
Negative
Individual Student Self-Concept
F
Red Shirting (Relative year-in-school effect)
Individual Student Math Achievement
Positive
Red Shirting (Relative Year in School)
Positive
Individual Student Math Self-Concept


G
Integration of BFLPE and Dimensional Comparison Theory
Individual Ach (L1)
School-average Ach
Individual Math Ach
Individual Verbal Ach
Individual Science Ach
Ind1: Positive
Ind2: Negative
Ind3: >Ind2 & < Ind1
Individual Student Math Self-Concept
SA1: Negative
SA2: Positive
SA3: >SA1 & < SA2
School-Average Math Ach
School-Average Verbal Ach
School-Average Science Ach

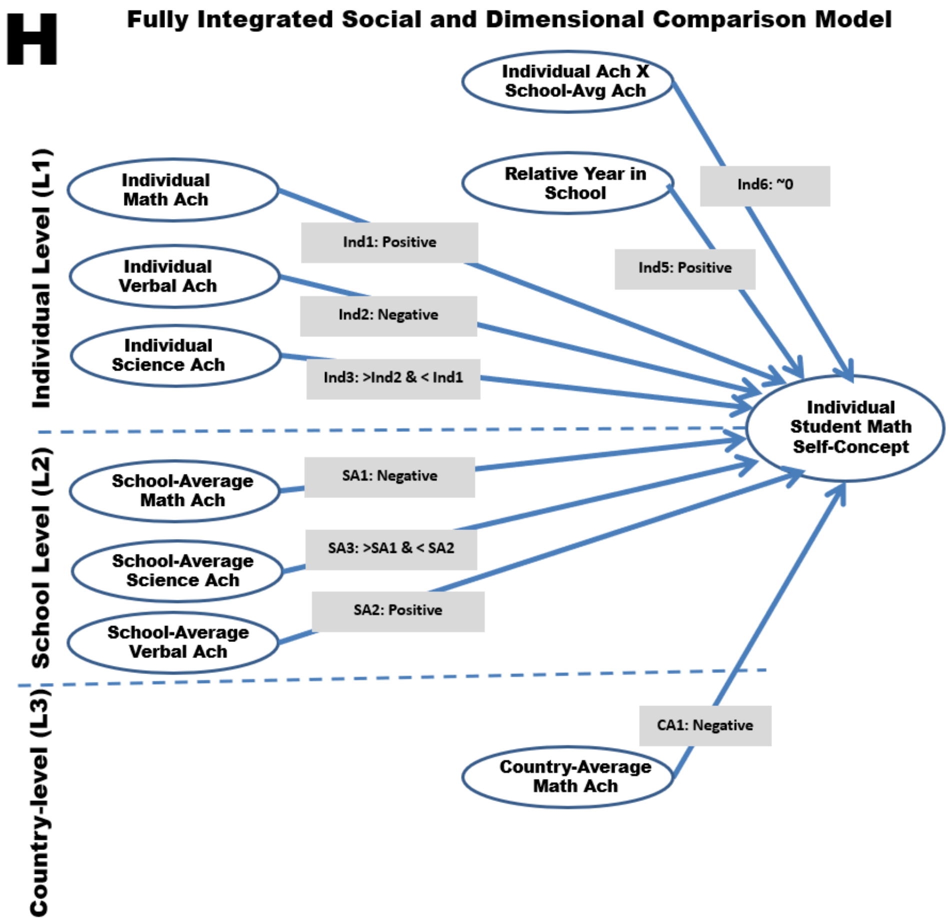

H
Fully Integrated Social and Dimensional Comparison Model
Individual Level (L1)
School Level (L2)
Country-level (L3)
Individual Ach X School-Avg Ach
Relative Year in School
Ind6: ~0
Ind5: Positive
Individual Math Ach
Individual Verbal Ach
Individual Science Ach
Ind1: Positive
Ind2: Negative
Ind3: >Ind2 & < Ind1
Individual Student Math Self-Concept
School-Average Math Ach
School-Average Science Ach
School-Average Verbal Ach
SA1: Negative
SA3: >SA1 & < SA2
SA2: Positive
CA1: Negative
Country-Average Math Ach

Figure 1. Individual and integrated frame-of-reference effects.

Note. (A) Internal-external frame-of-reference (I/E): Predicts that the effect on math self-concept is positive for math achievement but negative for verbal achievement (and that the effect on verbal self-concept is positive for verbal achievement but negative for math achievement).

(B) Dimensional comparison theory (DCT): predicts (the same as the I/E model) that the effect on math self-concept (MSC) is positive for math achievement but negative for verbal achievement. Extending the I/E model, DCT predicts that the effect of science achievement on MSC will be less negative (closer to zero) than the negative effect of reading achievement on MSC, and less positive (closer to zero) than the positive effects of math achievement on MSC

(C) Big-fish-little-pond effect (BFLPE): Predicts that the effect of school-average achievement on MSC is negative.

(D) "Bright student" hypothesis: Predicts that bright students will not suffer the BFLPE (i.e., that the individual by school-average achievement will be positive), whereas BFLPE theory predicts that the interaction will be close to zero.

(E) Paradoxical cross-cultural effect: predicts that the effect of country-average math achievement on MSC is negative.

(F) Relative year in school effect (red shirting): predicts the effect of being old for your year in school (starting late or repeating a year in school) on MSC is negative.

(G) Integration of BFLPE and dimensional comparison theory (Big-fish-little-pond compensatory effect, BFLPE-Compensatory effect): Predicts same effects as in the I/E and DCT models (Figure 1A and 1B) and the BFLPE model (1C). The new predictions are for the effects for school-average achievements. Integrating the logic of the DCT and BFLPE models in regard to school-average achievement, the effects are predicted to be in the opposite direction of those for individual achievement; negative for math achievement (SA1), positive for verbal achievement (SA2), and in between for science achievement (SA3 > SA1 & SA3 < SA2).

(H) Fully integrated social and dimensional comparison models. In this model, all effects presented in Figures 1A to 1G are integrated into a single three level model.

Table 1

Effect Sizes of the Big-Fish-Little-Pond Effect (BFLPE) Across Four PISA Data Collections

Country	Pisa 2012 Math	PISA 2006 Science	PISA 2003 Math	PISA 2000 General
Albania	-.23			
Argentina	-.49	-.18		
Australia	-.33	-.17	-.28	-.23
Austria	-.47	-.23	-.48	-.23
Azerbaijan		-.15		
Belgium	-.37	-.18	-.45	-.12
Brazil	-.48	-.12	-.37	-.26
Bulgaria	-.31	-.07		
Canada	-.44	-.23	-.43	
Chile	-.49	-.12		
Chinese Taipei	-.23	-.08		
Colombia	-.39	-.13		
Connecticut (USA)	-.25			
Costa Rica	-.32			
Croatia	-.28	-.12		
Czech Republic	-.45	-.22	-.45	-.24
Denmark	-.47	-.19	-.30	-.17
Estonia	-.42	-.18		
Finland	-.32	-.25	-.30	-.14
Florida (USA)	-.42			
France	-.43	-.23	-.38	
Germany	-.62	-.30	-.71	-.30
Greece	-.30	-.15	-.17	
Hong Kong-China	-.35	-.21	-.20	
Hungary	-.43	-.21	-.32	-.05
Iceland	-.37	-.17	-.21	-.18
Indonesia	-.24	-.20	-.24	
Ireland	-.34	-.19	-.10	-.24
Israel	-.44	-.22		
ITALY	-.36	-.21	-.41	-.36
Japan	-.29	-.10	-.31	
Jordan	-.38	-.11		
Kazakhstan	-.41			
Korea	-.14	.05	-.01	-.02
Kyrgyztan		-.19		
Latvia	-.34	-.12	-.22	-.06
Liechtenstein	-.39		-.55	-.20
Lithuania	-.41	-.14		
Luxembourg	-.29	-.08	-.43	-.17
Macao-China	-.39	-.16	-.33	

Table 1 Continued

Malaysia	-.35			
Massachusetts (USA)	-.37			
Mexico	-.38	-.06	-.36	-.08
Montenegro	-.35	-.14		
Netherlands	-.46	-.29	-.70	-.26
New Zealand	-.34	-.24	-.31	-.26
Norway	-.39	-.20	-.17	-.18
Russian Federation	-.24			
Peru	-.53			
Poland	-.40	-.13	-.28	
Portugal	-.30	-.27	-.21	-.18
Qatar	-.36	-.27		
Romania	-.16	-.09		
Russian Federation	-.33	-.22	-.19	-.21
Serbia	-.27	-.14	-.18	
Shanghai-China	-.24			
Singapore	-.19			
Slovak Republic	-.42	-.19	-.41	
Slovenia	-.51	-.19		
Spain	-.31	-.08	-.24	
Sweden	-.36	-.18	-.20	-.33
Switzerland	-.37	-.20	-.45	-.17
Thailand	-.31	-.18	-.19	
Tunisia	-.14	-.12	-.16	
Turkey	-.31	-.11	-.25	
United Arab Emirates	-.43			
United Kingdom	-.28	-.23	-.34	-.23
United States of America	-.40	-.35	-.23	-.26
Uruguay	-.37	-.16	-.24	
Viet Nam	-.13			
Mean	-.35	-.17	-.31	-.20

Note. The effect sizes are for four big-fish-little-pond effect studies based on PISA data collections between 2000 and 2012, which are reported in previously published results: PISA200 (Marsh & Hau, 2003: 103,558 students from 26 countries); PISA2003 (Seaton, Marsh & Craven, 2009, 2010: 265,180 students from 41 countries); PISA2006 (Nagengast & Marsh, 2012: 397,500 students from 57 countries); and PISA2012 (Marsh, Parker & Pekrun, 2017: 485,490 from 68 countries/regions), including the PISA2012 data considered in the present investigation.

Table 2

Integrated Big-Fish-Little-Pond and Internal/External Frame-of-reference Models

	Model 1		Model 2		Model 3	
Parameter	*β*	*SE*	*β*	*SE*	*β*	*SE*
Constant	-.06	.026	-.03	.027	-.05	.026
L1-Math Ach (linear)	.93	.032	.90	.032	.91	.034
L1-Math Ach(quad)	.10	.004	.10	.004	.10	.003
L2-Math Ach (BF LPE)	-.46	.071	-.47	.066	-.50	.027
L1 x L2 Math Ach Interaction	-.07	.004	-.06	.005	-.05	.015
L1-Relative Year in School	.05	.002	.05	.003	.06	.006
L3-Math Ach (country)	-.35	.051	-.34	.048	-.33	.052
L1-Read Ach	-.33	.015	-.28	.017	-.32	.023
L1- Science Ach	-.14	.016	-.16	.014	-.18	.016
L2-Science Ach	.02	.045	.03	.042	.06	.017
L2-Read Ach	.15	.029	.14	.028	.14	.024
Covariates						
Female			-.09	.008	-.07	.005
Immigrant (1st gen)			.13	.010	.14	.008
Immigrant (2ndt gen)			.08	.009	.09	.009
Parent Education			.03	.003	.03	.002
Parent Occupation			.00	.003	.00	.003
Random Effects						
L3-Res Var Intercept	.04	.007	.03	.006	.04	.008
L3-Res Var L1-Math Ach					.07	.013
L3-Res Var L2-Math Ach					.03	.007
L3-Res Var L1 x L2-Math Ach					.01	.002
L3-Res Var L1-Relative Year					.00	.000
L3-Res Var L1-Read Ach					.03	.006
L3-Res Var L1-Science Ach					.01	.002
L3-Res Var L2-Science Ach					.01	.003
L3-Res Var L2-Read Ach					.02	.005
L2-Res Var Intercept	.03	.002	.02	.001	.02	.001
L1-Res Var Intercept	.74	.008	.74	.007	.72	.003

Note. Predictor Variables: L1 = Student level; L2 = School level; L3 = Country level; Ach = achievement. Math Ach(quad) = Quadratic Component of individual math achievement. Res Var =

Residual variance components. Parameter estimates that differ from zero by more than two standard errors are statistically significant ($p < .05$). For these data, MSC was correlated with the three achievements (.26, math; .17, science, and .12, reading), whereas the three achievement scores were highly correlated with each other (.86 to .90).

Table 3*Parameter Estimates and Standard Errors (SE) for Each of 68 Countries/Regions*

	L2 Math Achieve		Relative Year in School		L1-Read Achieve		L2-Read Achieve	
Country/ Region	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
1 Albania	-.28	.08	.03	.02	.04	.02	.24	.06
2 United Arab Emirates	-.68	.05	.04	.01	-.27	.03	-.03	.05
3 Argentina	-.64	.08	.07	.01	-.20	.03	.29	.06
4 Australia	-.64	.06	.08	.01	-.51	.03	-.07	.05
5 Austria	-.44	.08	.07	.02	-.25	.04	.23	.06
6 Belgium	-.53	.06	.07	.01	-.42	.03	.05	.06
7 Bulgaria	-.32	.07	.02	.02	-.20	.03	.06	.06
8 Brazil	-.51	.05	.02	.01	-.29	.02	.12	.05
9 Canada	-.71	.04	.10	.01	-.29	.02	.28	.05
10 Switzerland	-.65	.06	.09	.01	-.63	.02	.07	.06
11 Chile	-.57	.08	.09	.01	-.23	.03	.03	.08
12 Colombia	-.45	.06	.05	.01	-.17	.03	.15	.06
13 Costa Rica	-.36	.08	.05	.01	-.25	.03	.03	.07
14 Czech Republic	-.51	.06	.07	.02	-.34	.03	.26	.06
15 Germany	-.82	.07	.07	.01	-.52	.04	.44	.07
16 Denmark	-.78	.08	.07	.02	-.68	.03	.28	.07
17 Spain	-.65	.04	.08	.01	-.29	.02	.07	.04
18 Estonia	-.56	.09	.08	.02	-.31	.04	.33	.08
19 Finland	-.68	.07	.18	.01	-.46	.02	.19	.06
20 France	-.52	.07	.04	.02	-.46	.03	.40	.06
21 United Kingdom	-.67	.07	.07	.01	-.69	.03	.13	.05
22 Greece	-.39	.08	.05	.02	-.25	.03	.34	.07
23 Hong Kong-China	-.65	.08	.06	.01	-.58	.04	.16	.08
24 Croatia	-.41	.08	.03	.02	-.31	.04	.15	.07
25 Hungary	-.54	.08	.06	.02	-.33	.04	-.02	.07
26 Indonesia	-.27	.07	.04	.01	-.06	.03	.13	.06
27 Ireland	-.60	.08	.06	.01	-.35	.04	.20	.07
28 Iceland	-.54	.10	.06	.03	-.39	.03	.10	.08
29 Israel	-.54	.08	.05	.02	-.27	.03	.03	.06
30 Italy	-.22	.04	-.01	.01	-.19	.02	-.03	.03
31 Jordan	-.32	.07	.05	.02	-.09	.03	.17	.05
32 Japan	-.52	.07	.06	.03	-.42	.03	-.12	.07
33 Kazakhstan	-.27	.06	.03	.01	.02	.03	.32	.06
34 Korea	-.48	.08	.06	.03	-.55	.04	.15	.08
35 Liechtenstein	-.54	.13	.07	.03	-.47	.10	.14	.12
36 Lithuania	-.51	.08	.10	.02	-.36	.04	.11	.07
37 Luxembourg	-.45	.11	.07	.01	-.46	.03	.13	.09
38 Latvia	-.51	.08	.06	.02	-.27	.04	.15	.07
39 Macao-China	-.52	.09	.08	.01	-.38	.03	.05	.09
40 Mexico	-.43	.04	.04	.01	-.21	.02	.10	.04
41 Montenegro	-.44	.11	.03	.02	-.28	.03	.23	.09
42 Malaysia	-.65	.07	.07	.03	-.17	.03	.16	.07
43 Netherlands	-.58	.08	.11	.02	-.48	.04	.17	.07
44 Norway	-.61	.09	.07	.03	-.29	.03	.20	.07
45 New Zealand	-.59	.08	.09	.02	-.46	.03	-.02	.07
46 Peru	-.60	.07	.05	.01	-.21	.03	.17	.07
47 Poland	-.72	.08	.11	.02	-.45	.03	.15	.07
48 Portugal	-.44	.08	.04	.01	-.30	.03	.01	.07
49 Qatar	-.47	.07	.03	.01	-.13	.02	.43	.06
50 Shanghai-China	-.44	.09	.11	.01	-.65	.04	.11	.09

Table 3 Continued

51 Perm (Russian Fed)	-.35	.10	.06	.02	-.11	.04	.20	.09
52 Florida (USA)	-.65	.11	.06	.02	-.37	.05	.20	.10
53 Connecticut (USA)	-.41	.11	.05	.02	-.31	.05	.20	.10
54 Massachusetts (USA)	-.51	.11	.08	.02	-.31	.05	.02	.10
55 Romania	-.15	.07	.05	.02	-.12	.03	.04	.06
56 Russian Federation	-.36	.07	.06	.01	-.13	.03	.18	.06
57 Singapore	-.36	.09	.02	.02	-.38	.04	-.03	.08
58 Serbia	-.27	.08	.06	.03	-.26	.03	.05	.07
59 Slovak Republic	-.45	.07	.06	.01	-.32	.03	.04	.06
60 Slovenia	-.54	.07	.06	.02	-.30	.04	.27	.05
61 Sweden	-.61	.09	.10	.03	-.36	.03	.23	.07
62 Chinese Taipei	-.41	.07	.06	.02	-.48	.04	.13	.08
63 Thailand	-.42	.06	.06	.01	-.29	.03	.16	.06
64 Tunisia	-.27	.08	.07	.02	-.20	.03	.01	.07
65 Turkey	-.29	.07	.04	.01	-.10	.04	.16	.07
66 Uruguay	-.55	.08	.06	.01	-.31	.03	.26	.07
67 United States of America	-.73	.09	.07	.02	-.42	.04	.10	.06
68 Vietnam	-.25	.07	.05	.02	-.15	.03	.24	.05

Note. Results based on the random effects in Model (Table 2) are presented separately for each of the 68 countries/regions. Parameter estimates that differ from zero by more than two standard errors are statistically significant ($p < .05$); those in grey are not statistically significant ($p > .05$). Estimates are standardized effects in which all independent and dependent variables are standardized.

Supplemental Materials

1 Supplemental Section 1. Red-shirting and Relative Year in School Effect.

2 Supplemental Section 2. PISA2012 Sampling Design and Implications for Multiple Imputation Analysis

3 Supplemental Section 3. Extended test of the I/E and dimensional comparison theory at the individual student level (Figures 1A and 1B)

- Supplemental Table 1**.** *Math Self-Concept Predictions: Extended Internal/External Frame-of-reference Models*

4 Supplemental Section 4. Extended versions of Tables 2 and 3 from Main Text

- Supplemental Table 2 (Extension of Table 2 from Main Text). *Parameter Estimates, Standard Errors (SE), Confidence Intervals and p-values for All Effects*
- Supplemental Table 3 (Extension of Table 3 from Main Text) Parameter Estimates, Standard Errors (SE), Confidence Intervals and p-values for Each of 68 Countries/Regions
- Supplemental Table 4. Parameter Estimates and Standard Errors (SE) for Each of 68 Countries/Regions (An Extension of Table 3 in the Main Text)

5 Supplemental Section 5. A Universalist Perspective on Cross-Cultural Generalizability and Variability(Hypothesis 4)

6 Supplemental Section 6. Assumptions of Causality and Underlying Processes.

7 Supplemental Section 7. MLWIN Macro used to fit Model 1 in Table 2.

8 Supplemental References. References for articles cited in Supplemental Materials that are not also cited in the main text.

Supplemental Section 1.

Red-shirting and Relative Year in School Effect.

The term “red shirting” came from US college sports and referred to holding a college-athlete back a year to develop skills and extend eligibility. By way of analogy, in school settings the term refers to holding students back so that they are relatively older compared to their classmates. Research (e.g., Spitzer, Cupp, and Parke, 1995; also see Gladwell, 2008; Marsh, 2016; Marsh, Pekrun, et al., 2017; Parker, Marsh, Thoemmes & Biddle, 2019) showed that this had positive effects on achievement and multiple components of self-concept, although there is controversy as to whether the effects are short- or long-term.

Integrating this red shirting phenomena into academic self-concept research, Marsh and colleagues (Marsh, 2016; Marsh, Parker & Pekrun, 2018; Marsh, Pekrun et al., 2017) noted that students who are older than their classmates tend to be physically, emotionally, and academically more mature. In this respect, classmates who are younger and less mature are likely to provide a less demanding frame-of-reference. For these reasons, it was predicted and found that the effect of relative age on academic self-concept (red shirting) was positive. Marsh (2016; also see Marsh, Parker & Pekrun, 2018) distinguished between relative year in school associated with starting school at an older age (i.e., red shirting) and with having repeated a year in school. However, the results were essentially the same, and neither of these variables had an effect (positive or negative) beyond the effect of red shirting. On this basis, they argued that relative year in school was the critical variable, rather than the reason why a student was relatively old in relation to his/her year in school.

In subsequent research, Marsh, Pekrun, et al. (2018) focused more specifically on student retention (repeating a year in school) in relation to a set of 10 outcome variables (e.g., math self-concept, self-efficacy, anxiety, relations with teachers, parents and peers, school grades, and standardized achievement test scores). Consistent with a priori predictions based on extensions of the BFLPE and red shirting, they found largely positive effects associated with repeating a year in school. The strong design was based on a large, representative sample of German students (N = 1,325, *M* age = 11.75 at time 1 at the end of primary school) who were tested each year during the first 5 years of secondary school. The design featured 4 independent retention groups (different groups of students, each repeating one of the first four years of secondary school), with multiple posttest waves to evaluate short- and long-term effects, controlling for covariates (gender, age, socioeconomic status, primary school grades, and IQ) and one or more sets of 10 outcomes collected prior to retention. Tests of developmental invariance demonstrated that the effects of retention (controlling for covariates and pre-retention outcomes) were highly consistent across this potentially volatile early to middle adolescent period; largely positive effects in the first year following retention were maintained in subsequent school years following retention. The largest effect of retention (approximately 1 *SD*) was a positive effect on math school grades in the first year following retention. Because school grades are a primary driver of academic self-concept, it is not surprising that MSCs increased as well. Importantly, Marsh, Pekrun, et al. were able to show that these positive results in the first year following retention were largely maintained in subsequent school years. Marsh, Pekrun et al. also noted that the results were consistent over different cohorts of students. Particularly considering that these results are contrary to at least some of the accepted wisdom about school retention, the findings have important implications for educational researchers, policymakers, and parents.

Marsh, Pekrun, et al. (2018) also noted that their results were consistent with theoretical perspective on mastery learning and Matthew effects. There is clear evidence from mastery learning interventions that weaker students might merely need more time to master new material, material that can be mastered more quickly by stronger students. If weak students are given sufficient time and resources to achieve mastery, the differences between more and less able students will diminish and achieving mastery has potentially profound effects on positive self-beliefs and motivations to learn. Relatedly, there is evidence from studies of the Matthew effect showing that without intervention, students who fall behind at any particular stage in schooling tend to fall behind even further in subsequent school years. Hence, early intervention is critical to break the vicious cycle created by Matthew effects. Consistent with these theoretical and empirical perspectives, the fact that retained students had an extra year to learn the materials that had led to their retention not only helped them to learn those materials more effectively in the first year following retention, but also resulted in more positive self-beliefs and gave them a stronger basis for learning new materials in subsequent school years. Hence, retention can be seen as a potentially useful intervention to counter the negative consequences of failure to learn critical academic materials. We also note that retained students tend to be more mature (i.e., a year older than their new classmates following retention). Indeed, it is curious that there seems to be widespread support for holding students back when they start school so that they are among the oldest in their class, rather than the youngest, but the opposite view prevails in terms of holding students back by repeating a school year when they have not adequately mastered the materials. However, Marsh (2016) argues that the advantage of being relatively older than classmates in terms of academic self-concept is similar for students who started late and those who repeat a year in school, and that this pattern of results has broad cross-cultural generalizability. Our results are consistent with those conclusions, but extend them in important new directions—particularly in relation to academic achievement and the long-term maintenance of short-term benefits of retention.

It is also relevant to note that repeating a year in school is a single, even if traumatic, experience; while comparison with peers is a daily experience. Thus, the repeated exposure to a frame-of-reference has a much larger effect than single discreet events. However, in the Marsh, Pekrun et al. (2017) study, data were collected shortly before the end of each school year – before students formally knew that they would be asked to repeat a grade in that year, and almost a full year following the time that previously retained students had repeated a grade. In the present investigation we have not differentiated between students are older because they were older when they first started school (the more typical reason) or because they had been held back and repeated a grade. However, the Marsh (2016) study suggests that the outcomes are the similar for both groups, and that relative age explains the results for both reasons.

Supplemental Section 2:

PISA2012 Sampling Design and Implications for Multiple Imputation Analysis

In the PISA2012 data, math achievement at the individual student level is represented by five plausible values (Mislevy, Beaton, Kaplan & Sheehan, 1992; von Davier, Gonzalez & Mislevy, 2009) that represent the individual ability distribution of the student, and facilitate secondary analysis. There were no missing data for the plausible values of math achievement. However, correct analysis of plausible values requires that all models are run separately for each plausible value and the results integrated using principles of multiple imputation analysis (see OECD, 2014).

As noted in the PISA2012 Technical Report (OECD, 2014), whereas rotation of cognitive test items has been used regularly in PISA data collections, 2012 is the first time this strategy has been used for student context surveys. This was done to increase the content coverage, whilst maintaining the amount of time needed to complete the survey. There were three survey forms, each of which contained a common set of items and a rotated section. In the rotated section, students completed 2/3 of the rotated items, such that allocation was based on the use of intact scales that were balanced in terms of correlations with performance. Because responses based on this strategy were purely missing completely at random, they were appropriately handled using multiple imputation. Importantly, year in school and related variables on repeating grade, school starting age, and current student age were in the common part of the survey, so that there were few missing values for any of these variables. However, math self-concept was rotated, so that approximately 1/3 of these variables were missing by design. For the present purposes, these missing values were appropriately handled by the multiple imputation strategy.

Because of the design of the PISA2012—the provision of five plausible values to represent achievement and missing by design for survey items—we used multiple imputation (Rubin, 1987) to deal with the missing data. Using a large imputation model, five imputed datasets were created using MCMC imputation, including dummy variables to represent the 68 countries. The decision to use multiple imputation was based on the need to include many auxiliary variables in the imputation model, the need to use plausible values for the achievement test that require an analytical strategy akin to the analysis of multiply imputed datasets, and also the rotation strategy first introduced in PISA2012 for the student survey, with non-cognitive items. To account for the five plausible values of achievement provided as part of the PISA database, each of the five imputations included one of these five plausible values, and results from the five imputations were combined using Rubin's (1987) approach.

Supplemental Section 3:

Extended test of the I/E and dimensional comparison theory at the individual student level (Figures 1A & 1B)

In the main text, tests of Hypothesis 1 are presented for Model 3 that incorporated both individual student achievements and school-average achievements. Although clearly appropriate, it leaves open the question as to how the results for the DCT predictions (Hypothesis 1) would differ if the school-average measures of achievement were not included in the model. Hence, here we provide an evaluation of support for the I/E model and DCT (without school-average measures of achievement).

The effect of math achievement on MSC, not surprisingly, was substantial and positive (β = .45, Model 1.1). In keeping with the classic I/E model, when both math and reading are used to predict MSC, the effect of math achievement increased (β = .82, Model 1.2) and the effect of reading achievement was negative (β = -.45). When all three achievement scores were used to predict MSC (Model 1.3), the effect of math achievement increased even further (β = .91), whereas the effects of both reading and science were negative. Consistent with dimensional comparison theory, the effect of science achievement (β = -.16, *SE* = .012) is less negative (closer to zero) than the effect of reading achievement (β = -.39, *SE* = .020), but less positive (closer to zero) than the positive effect of L1-math achievement. In summary there is good support for Hypotheses 4A and 4B when school-average measures of achievement are not included. Furthermore, these results are highly similar to those in the main text for which the school-average measures are included.

Supplemental Table 1

Math Self-Concept Predictions: Extended Internal/External Frame-of-reference Models

	Model 1.1		Model 1.2		Model 1.3		Model 1.4	
Parameters	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Fixed Effects								
Math Achievement	.45	.016	.82	.030	.91	.033	.79	.029
Read Achievement			-.45	.021	-.39	.020		
Science Achievement					-.16	.012	-.38	.019
Random Effects								
L3 Res Var Intercept	.11	.019	.11	.019	.11	.019	.11	.018
L3 Res Var Math Ach	.02	.003	.06	.010	.07	.012	.06	.010
L3 Res Var Read Ach			.03	.005	.03	.005		
L3 Res Var Science Ach					.01	.002	.02	.004
L2 Res Var Intercept	.05	.001	.04	.001	.04	.001	.05	.001
L1 Res Var Intercept	.79	.002	.74	.002	.74	.002	.76	.002

Note. Est = parameter estimates; Res Var = Residual variance components; L1 = Student level; L2 = School-average level; L3 = Country-average level.

Supplemental Section 5.

A Universalist Perspective on Cross-Cultural Generalizability and Variability (Hypothesis 4)

Cross-cultural generalizability. In the Universalist perspective in cross-cultural psychology, there is an emphasis on replicability of results, empirical tests, and cross-cultural generalizability of support for theoretical predictions. From a Universalist perspective, the remarkable cross-cultural support for all five frame-of-reference effects suggests that the meta-theoretical model upon which they are based constitutes a cross-culturally valid theory of ASC formation.

In support of this interpretation of cross-cultural generalizability, it is relevant to juxtapose cross-cultural research and meta-analysis as bases for establishing a Universalist perspective on generalizability. The two approaches share many of the same strengths and weaknesses, but there are some important differences. Meta-analysis is broadly seen as the gold standard of good science, testing the generalizability and replicability of support for predictions across a potential universe of related studies. Historically, the focus of meta-analysis has been on the overall effect size, based on the sample of studies being considered. However, this perspective is potentially overly-narrow, and has been broadened to include a focus on multi-level modelling, which places more emphasis on study-to-study variation in effect sizes, and on generalizability to a broader population of potential studies. Thus, both for meta-analysis and for cross-cultural research from a Universalist perspective, it is important to consider simultaneously the overall effects (average effects across individual studies in meta-analysis, or across individual countries in cross-cultural research) and generalizability-based residual variation (study-to-study variation in meta-analyses, and country-to-country variation in cross-cultural research). For each of these approaches to generalizability, evolving multi-level approaches are important in terms of juxtaposing fixed effects (overall effect sizes), and random effects (evidence related to the generalizability of the effects).

Meta-analysis, despite its many strengths, has important limitations. In particular, it suffers in terms of the heterogeneity of research in respect to materials, participants, measures, and research designs used in the diverse samples of studies considered. Individual studies included in meta-analyses typically are based on samples of convenience, even when there is random assignment to conditions for participants within each study. Further, particularly in psychological research, the sample of studies considered in a typical meta-analysis is likely to be biased in terms of the participants (over-representation of middle-class participants from Western countries, especially university students) and other influences, such as publication bias (under-representation of non-significant effects) and over-representation of studies published in English-language journals). Importantly, meta-analysts are limited to what is actually published and reported in existing studies. Particularly from a cross-cultural perspective, meta-analyses are unlikely to represent a Universalist perspective in relation to representative coverage of countries around the world.

Relatedly, because meta-analysts rarely have access to individual level data, this substantially curtails their ability to test hypotheses about how effects are related to individual participant characteristics. Thus, for example, a meta-analysis of BFLPE studies would not be able to test the “bright student” (Hypothesis 2B) prediction that the overall size of the effect varies as a function of individual achievement, because the analyst typically would not have access to data at the individual student level. Cross-cultural research likewise has related weaknesses, particularly in terms of the comparability of samples and materials across countries, the representativeness of samples from different countries, and the typically small numbers of countries contained in any one cross-cultural study.

Hussong, Curran, and Bauer (2013) similarly noted many of these limitations to meta-analysis and proposed what they referred to as integrative data analysis (IDA), which is based on the integration of primary data from different studies. Because IDA uses individual level data it is possible to test many hypotheses that could not be tested with traditional meta-analysis data (e.g., the interaction between L1-student and L2-school-average achievement). Although this overcomes some of the problems with traditional meta-analysis, there are important obstacles to IDA, such as accounting for sampling heterogeneity across studies, harmonization of incompatible measures, data linkage problems, and obtaining sufficient data sets to generalize from the samples back to the population of interest. In particular, both IDA and traditional meta-analysis studies are likely to be weak in terms having representative samples from a sufficient number of countries to test cross-cultural generalizability that is a focus of Universalist studies.

Large-scale cross-cultural PISA studies are stronger than typical cross-cultural and meta-analyses, and even IDAs. Many of the problems of traditional meta-analysis and cross-cultural research are overcome: there are nationally representative samples from many different countries; the constructs are psychometrically strong and have been pilot tested for cross-cultural appropriateness; all the participants receive the same materials; all data is available at the level of the individual participant. Although such PISA analyses could be seen as a special case IDA approach, they also overcome most of the obstacles to the typical IDA study—particularly in relation to tests of the universality of cross-cultural effects. There are, of course, limitations in the use of PISA data—for example, the age of participants and the cross-sectional nature of the data (see subsequent discussion of limitations), as well as the fact that not all countries are included in PISA data collections. Nevertheless, PISA data provides a potentially stronger basis for evaluating the cross-cultural generalizability of theoretical predictions than do traditional cross-cultural comparisons, meta-analyses, or even IDAs.

Supplemental Section 4:

Supplemental Table 2 (Extension of Table 2 from Main Text)

Parameter Estimates, Standard Errors (SE), Confidence Intervals and p-values for All Effects

Parameter	Model 1						Model 2						Model 3					
	β	*SE*	*t*	CI-Lo	CI-Hi	*p*	*β*	*SE*	*t*	CI-Lo	CI-Hi	*p*	*β*	*SE*	*t*	CI-Lo	CI-Hi	*p*
Constant	-.06	.03	-2.31	-.01	-.11	.021	-.03	.03	-1.11	.02	-.08	.267	-.05	.03	-1.92	.00	-.10	.054
L1-Math Ach (linear)	.93	.03	29.06	.99	.87	.000	.90	.03	28.13	.96	.84	.000	.91	.03	26.76	.98	.84	.000
L1-Math Ach(quad)	.10	.00	25.00	.11	.09	.000	.10	.00	25.00	.11	.09	.000	.10	.00	33.33	.11	.09	.000
L2-Math Ach (BF LPE)	-.46	.07	-6.48	-.32	-.60	.000	-.47	.07	-7.12	-.34	-.60	.000	-.50	.03	-18.52	-.45	-.55	.000
L1 x L2 Math Ach Interaction	-.07	.00	-17.50	-.06	-.08	.000	-.06	.01	-12.00	-.05	-.07	.000	-.05	.02	-3.33	-.02	-.08	.001
L1-Relative Year in School	.05	.00	25.00	.05	.05	.000	.05	.00	16.67	.06	.04	.000	.06	.01	1.00	.07	.05	.000
L3-Math Ach (country)	-.35	.05	-6.86	-.25	-.45	.000	-.34	.05	-7.08	-.25	-.43	.000	-.33	.05	-6.35	-.23	-.43	.000
L1-Read Ach	-.33	.02	-22.00	-.30	-.36	.000	-.28	.02	-16.47	-.25	-.31	.000	-.32	.02	-13.91	-.27	-.37	.000
L1- Science Ach	-.14	.02	-8.75	-.11	-.17	.000	-.16	.01	-11.43	-.13	-.19	.000	-.18	.02	-11.25	-.15	-.21	.000
L2-Science Ach	.02	.05	.44	.11	-.07	.657	.03	.04	.71	.11	-.05	.475	.06	.02	3.53	.09	.03	.000
L2-Read Ach	.15	.03	5.17	.21	.09	.000	.14	.03	5.00	.19	.09	.000	.14	.02	5.83	.19	.09	.000
Covariates																		
Female							-.09	.01	-11.25	-.07	-.11	.000	-.07	.01	-14.00	-.06	-.08	.000
Immigrant (1st gen)							.13	.01	13.00	.15	.11	.000	.14	.01	17.50	.16	.12	.000
Immigrant (2ndt gen)							.08	.01	8.89	.10	.06	.000	.09	.01	1.00	.11	.07	.000
Parent Education							.03	.00	1.00	.04	.02	.000	.03	.00	15.00	.03	.03	.000
Parent Occupation							.00	.00	.00	.01	-.01	1.000	.00	.00	.00	---	---	---
Random Effects																		
L3-Res Var Intercept	.04	.01	5.71	.05	.03	.000	.03	.01	5.00	.04	.02	.000	.04	.01	5.00	.06	.02	.000
L3-Res Var L1-Math Ach													.07	.01	5.38	.10	.04	.000
L3-Res Var L2-Math Ach													.03	.01	4.29	.04	.02	.000
L3-Res Var L1 x L2-Math Ach													.01	.00	5.00	.01	.01	.000
L3-Res Var L1-Relative Year													.00	.00	.00	---	---	----
L3-Res Var L1-Read Ach													.03	.01	5.00	.04	.02	.000
L3-Res Var L1-Science Ach													.01	.00	5.00	.01	.01	.000
L3-Res Var L2-Science Ach													.01	.00	3.33	.02	.00	.001
L3-Res Var L2-Read Ach													.02	.01	4.00	.03	.01	.000
L2-Res Var Intercept	.03	.00	15.00	.03	.03	.000	.02	.00	2.00	.02	.02	.000	.02	.00	2.00	.02	.02	.000
L1-Res Var Intercept	.74	.01	92.50	.76	.72	.000	.74	.01	105.71	.75	.73	.000	.72	.00	24.00	.73	.71	.000

Note. Predictor Variables: L1 = Student level; L2 = School level; L3 = Country level; Ach = achievement. Math Ach(quad) = Quadratic Component of individual math achievement. Res Var = Residual variance components. For these data, MSC was correlated with the three achievements (.26, math; .17, science, and .12, reading), whereas the three achievement scores were highly correlated with each other (.86 to .90). Models 1-3 presented here are the same as in Table 2 in the main text, but have been expanded to include t-values, confidence intervals, and p-values.

Supplemental Table 3 (Extension of Table 3 from Main Text)

Parameter Estimates, Standard Errors (SE), Confidence Intervals and p-values for Each of 68 Countries/Regions

Country/ Region	L2 Math Achieve						Relative Year in School						L1-Read Achieve						L2-Read Achieve					
	β	*SE*	*t*	CI-Lo	CI-Hi	*p*	β	*SE*	*t*	CI-Lo	CI-Hi	*p*	β	*SE*	*t*	CI-Lo	CI-Hi	*p*	β	*SE*	*t*	CI-Lo	CI-Hi	*p*
1 Albania	-.28	.08	-3.50	-.12	-.44	.000	.03	.02	1.50	.07	-.01	.134	.04	.02	2.00	.08	.00	.046	.24	.06	4.00	.36	.12	.000
2 United Arab Emirates	-.68	.05	-13.60	-.58	-.78	.000	.04	.01	4.00	.06	.02	.000	-.27	.03	-9.00	-.21	-.33	.000	-.03	.05	-.60	.07	-.13	.549
3 Argentina	-.64	.08	-8.00	-.48	-.80	.000	.07	.01	7.00	.09	.05	.000	-.20	.03	-6.67	-.14	-.26	.000	.29	.06	4.83	.41	.17	.000
4 Australia	-.64	.06	-1.67	-.52	-.76	.000	.08	.01	8.00	.10	.06	.000	-.51	.03	-17.00	-.45	-.57	.000	-.07	.05	-1.40	.03	-.17	.162
5 Austria	-.44	.08	-5.50	-.28	-.60	.000	.07	.02	3.50	.11	.03	.000	-.25	.04	-6.25	-.17	-.33	.000	.23	.06	3.83	.35	.11	.000
6 Belgium	-.53	.06	-8.83	-.41	-.65	.000	.07	.01	7.00	.09	.05	.000	-.42	.03	-14.00	-.36	-.48	.000	.05	.06	.83	.17	-.07	.405
7 Bulgaria	-.32	.07	-4.57	-.18	-.46	.000	.02	.02	1.00	.06	-.02	.317	-.20	.03	-6.67	-.14	-.26	.000	.06	.06	1.00	.18	-.06	.317
8 Brazil	-.51	.05	-1.20	-.41	-.61	.000	.02	.01	2.00	.04	.00	.046	-.29	.02	-14.50	-.25	-.33	.000	.12	.05	2.40	.22	.02	.016
9 Canada	-.71	.04	-17.75	-.63	-.79	.000	.10	.01	1.00	.12	.08	.000	-.29	.02	-14.50	-.25	-.33	.000	.28	.05	5.60	.38	.18	.000
10 Switzerland	-.65	.06	-1.83	-.53	-.77	.000	.09	.01	9.00	.11	.07	.000	-.63	.02	-31.50	-.59	-.67	.000	.07	.06	1.17	.19	-.05	.243
11 Chile	-.57	.08	-7.13	-.41	-.73	.000	.09	.01	9.00	.11	.07	.000	-.23	.03	-7.67	-.17	-.29	.000	.03	.08	.38	.19	-.13	.708
12 Colombia	-.45	.06	-7.50	-.33	-.57	.000	.05	.01	5.00	.07	.03	.000	-.17	.03	-5.67	-.11	-.23	.000	.15	.06	2.50	.27	.03	.012
13 Costa Rica	-.36	.08	-4.50	-.20	-.52	.000	.05	.01	5.00	.07	.03	.000	-.25	.03	-8.33	-.19	-.31	.000	.03	.07	.43	.17	-.11	.668
14 Czech Republic	-.51	.06	-8.50	-.39	-.63	.000	.07	.02	3.50	.11	.03	.000	-.34	.03	-11.33	-.28	-.40	.000	.26	.06	4.33	.38	.14	.000
15 Germany	-.82	.07	-11.71	-.68	-.96	.000	.07	.01	7.00	.09	.05	.000	-.52	.04	-13.00	-.44	-.60	.000	.44	.07	6.29	.58	.30	.000
16 Denmark	-.78	.08	-9.75	-.62	-.94	.000	.07	.02	3.50	.11	.03	.000	-.68	.03	-22.67	-.62	-.74	.000	.28	.07	4.00	.42	.14	.000
17 Spain	-.65	.04	-16.25	-.57	-.73	.000	.08	.01	8.00	.10	.06	.000	-.29	.02	-14.50	-.25	-.33	.000	.07	.04	1.75	.15	-.01	.080
18 Estonia	-.56	.09	-6.22	-.38	-.74	.000	.08	.02	4.00	.12	.04	.000	-.31	.04	-7.75	-.23	-.39	.000	.33	.08	4.13	.49	.17	.000
19 Finland	-.68	.07	-9.71	-.54	-.82	.000	.18	.01	18.00	.20	.16	.000	-.46	.02	-23.00	-.42	-.50	.000	.19	.06	3.17	.31	.07	.002
20 France	-.52	.07	-7.43	-.38	-.66	.000	.04	.02	2.00	.08	.00	.046	-.46	.03	-15.33	-.40	-.52	.000	.40	.06	6.67	.52	.28	.000
21 United Kingdom	-.67	.07	-9.57	-.53	-.81	.000	.07	.01	7.00	.09	.05	.000	-.69	.03	-23.00	-.63	-.75	.000	.13	.05	2.60	.23	.03	.009
22 Greece	-.39	.08	-4.88	-.23	-.55	.000	.05	.02	2.50	.09	.01	.012	-.25	.03	-8.33	-.19	-.31	.000	.34	.07	4.86	.48	.20	.000
23 Hong Kong-China	-.65	.08	-8.13	-.49	-.81	.000	.06	.01	6.00	.08	.04	.000	-.58	.04	-14.50	-.50	-.66	.000	.16	.08	2.00	.32	.00	.046
24 Croatia	-.41	.08	-5.13	-.25	-.57	.000	.03	.02	1.50	.07	-.01	.134	-.31	.04	-7.75	-.23	-.39	.000	.15	.07	2.14	.29	.01	.032
25 Hungary	-.54	.08	-6.75	-.38	-.70	.000	.06	.02	3.00	.10	.02	.003	-.33	.04	-8.25	-.25	-.41	.000	-.02	.07	-.29	.12	-.16	.775
26 Indonesia	-.27	.07	-3.86	-.13	-.41	.000	.04	.01	4.00	.06	.02	.000	-.06	.03	-2.00	.00	-.12	.046	.13	.06	2.17	.25	.01	.030
27 Ireland	-.60	.08	-7.50	-.44	-.76	.000	.06	.01	6.00	.08	.04	.000	-.35	.04	-8.75	-.27	-.43	.000	.20	.07	2.86	.34	.06	.004
28 Iceland	-.54	.10	-5.40	-.34	-.74	.000	.06	.03	2.00	.12	.00	.046	-.39	.03	-13.00	-.33	-.45	.000	.10	.08	1.25	.26	-.06	.211
29 Israel	-.54	.08	-6.75	-.38	-.70	.000	.05	.02	2.50	.09	.01	.012	-.27	.03	-9.00	-.21	-.33	.000	.03	.06	.50	.15	-.09	.617
30 Italy	-.22	.04	-5.50	-.14	-.30	.000	-.01	.01	-1.00	.01	-.03	.317	-.19	.02	-9.50	-.15	-.23	.000	-.03	.03	-1.00	.03	-.09	.317
31 Jordan	-.32	.07	-4.57	-.18	-.46	.000	.05	.02	2.50	.09	.01	.012	-.09	.03	-3.00	-.03	-.15	.003	.17	.05	3.40	.27	.07	.001
32 Japan	-.52	.07	-7.43	-.38	-.66	.000	.06	.03	2.00	.12	.00	.046	-.42	.03	-14.00	-.36	-.48	.000	-.12	.07	-1.71	.02	-.26	.086
33 Kazakhstan	-.27	.06	-4.50	-.15	-.39	.000	.03	.01	3.00	.05	.01	.003	.02	.03	.67	.08	-.04	.505	.32	.06	5.33	.44	.20	.000
34 Korea	-.48	.08	-6.00	-.32	-.64	.000	.06	.03	2.00	.12	.00	.046	-.55	.04	-13.75	-.47	-.63	.000	.15	.08	1.88	.31	-.01	.061
35 Liechtenstein	-.54	.13	-4.15	-.29	-.79	.000	.07	.03	2.33	.13	.01	.020	-.47	.10	-4.70	-.27	-.67	.000	.14	.12	1.17	.38	-.10	.243
36 Lithuania	-.51	.08	-6.38	-.35	-.67	.000	.10	.02	5.00	.14	.06	.000	-.36	.04	-9.00	-.28	-.44	.000	.11	.07	1.57	.25	-.03	.116
37 Luxembourg	-.45	.11	-4.09	-.23	-.67	.000	.07	.01	7.00	.09	.05	.000	-.46	.03	-15.33	-.40	-.52	.000	.13	.09	1.44	.31	-.05	.149
38 Latvia	-.51	.08	-6.38	-.35	-.67	.000	.06	.02	3.00	.10	.02	.003	-.27	.04	-6.75	-.19	-.35	.000	.15	.07	2.14	.29	.01	.032
39 Macao-China	-.52	.09	-5.78	-.34	-.70	.000	.08	.01	8.00	.10	.06	.000	-.38	.03	-12.67	-.32	-.44	.000	.05	.09	.56	.23	-.13	.579
40 Mexico	-.43	.04	-1.75	-.35	-.51	.000	.04	.01	4.00	.06	.02	.000	-.21	.02	-1.50	-.17	-.25	.000	.10	.04	2.50	.18	.02	.012
41 Montenegro	-.44	.11	-4.00	-.22	-.66	.000	.03	.02	1.50	.07	-.01	.134	-.28	.03	-9.33	-.22	-.34	.000	.23	.09	2.56	.41	.05	.011

42 Malaysia	-.65	.07	-9.29	-.51	-.79	.000	.07	.03	2.33	.13	.01	.020	-.17	.03	-5.67	-.11	-.23	.000	.16	.07	2.29	.30	.02	.022
43 Netherlands	-.58	.08	-7.25	-.42	-.74	.000	.11	.02	5.50	.15	.07	.000	-.48	.04	-12.00	-.40	-.56	.000	.17	.07	2.43	.31	.03	.015
44 Norway	-.61	.09	-6.78	-.43	-.79	.000	.07	.03	2.33	.13	.01	.020	-.29	.03	-9.67	-.23	-.35	.000	.20	.07	2.86	.34	.06	.004
45 New Zealand	-.59	.08	-7.38	-.43	-.75	.000	.09	.02	4.50	.13	.05	.000	-.46	.03	-15.33	-.40	-.52	.000	-.02	.07	-.29	.12	-.16	.775
46 Peru	-.60	.07	-8.57	-.46	-.74	.000	.05	.01	5.00	.07	.03	.000	-.21	.03	-7.00	-.15	-.27	.000	.17	.07	2.43	.31	.03	.015
47 Poland	-.72	.08	-9.00	-.56	-.88	.000	.11	.02	5.50	.15	.07	.000	-.45	.03	-15.00	-.39	-.51	.000	.15	.07	2.14	.29	.01	.032
48 Portugal	-.44	.08	-5.50	-.28	-.60	.000	.04	.01	4.00	.06	.02	.000	-.30	.03	-1.00	-.24	-.36	.000	.01	.07	.14	.15	-.13	.886
49 Qatar	-.47	.07	-6.71	-.33	-.61	.000	.03	.01	3.00	.05	.01	.003	-.13	.02	-6.50	-.09	-.17	.000	.43	.06	7.17	.55	.31	.000
50 Shanghai-China	-.44	.09	-4.89	-.26	-.62	.000	.11	.01	11.00	.13	.09	.000	-.65	.04	-16.25	-.57	-.73	.000	.11	.09	1.22	.29	-.07	.222
51 Perm (Russian Fed)	-.35	.1	-3.50	-.15	-.55	.000	.06	.02	3.00	.10	.02	.003	-.11	.04	-2.75	-.03	-.19	.006	.2	.09	2.22	.38	.02	.026
52 Florida (USA)	-.65	.11	-5.91	-.43	-.87	.000	.06	.02	3.00	.10	.02	.003	-.37	.05	-7.40	-.27	-.47	.000	.2	.1	2.00	.40	.00	.046
53 Connecticut (USA)	-.41	.11	-3.73	-.19	-.63	.000	.05	.02	2.50	.09	.01	.012	-.31	.05	-6.20	-.21	-.41	.000	.2	.1	2.00	.40	.00	.046
54 Massachusetts (USA)	-.51	.11	-4.64	-.29	-.73	.000	.08	.02	4.00	.12	.04	.000	-.31	.05	-6.20	-.21	-.41	.000	.02	.1	.20	.22	-.18	.841
55 Romania	-.15	.07	-2.14	-.01	-.29	.032	.05	.02	2.50	.09	.01	.012	-.12	.03	-4.00	-.06	-.18	.000	.04	.06	.67	.16	-.08	.505
56 Russian Federation	-.36	.07	-5.14	-.22	-.50	.000	.06	.01	6.00	.08	.04	.000	-.13	.03	-4.33	-.07	-.19	.000	.18	.06	3.00	.30	.06	.003
57 Singapore	-.36	.09	-4.00	-.18	-.54	.000	.02	.02	1.00	.06	-.02	.317	-.38	.04	-9.50	-.30	-.46	.000	-.03	.08	-.38	.13	-.19	.708
58 Serbia	-.27	.08	-3.38	-.11	-.43	.001	.06	.03	2.00	.12	.00	.046	-.26	.03	-8.67	-.20	-.32	.000	.05	.07	.71	.19	-.09	.475
59 Slovak Republic	-.45	.07	-6.43	-.31	-.59	.000	.06	.01	6.00	.08	.04	.000	-.32	.03	-1.67	-.26	-.38	.000	.04	.06	.67	.16	-.08	.505
60 Slovenia	-.54	.07	-7.71	-.40	-.68	.000	.06	.02	3.00	.10	.02	.003	-.3	.04	-7.50	-.22	-.38	.000	.27	.05	5.40	.37	.17	.000
61 Sweden	-.61	.09	-6.78	-.43	-.79	.000	.1	.03	3.33	.16	.04	.001	-.36	.03	-12.00	-.30	-.42	.000	.23	.07	3.29	.37	.09	.001
62 Chinese Taipei	-.41	.07	-5.86	-.27	-.55	.000	.06	.02	3.00	.10	.02	.003	-.48	.04	-12.00	-.40	-.56	.000	.13	.08	1.63	.29	-.03	.104
63 Thailand	-.42	.06	-7.00	-.30	-.54	.000	.06	.01	6.00	.08	.04	.000	-.29	.03	-9.67	-.23	-.35	.000	.16	.06	2.67	.28	.04	.008
64 Tunisia	-.27	.08	-3.38	-.11	-.43	.001	.07	.02	3.50	.11	.03	.000	-.2	.03	-6.67	-.14	-.26	.000	.01	.07	.14	.15	-.13	.886
65 Turkey	-.29	.07	-4.14	-.15	-.43	.000	.04	.01	4.00	.06	.02	.000	-.1	.04	-2.50	-.02	-.18	.012	.16	.07	2.29	.30	.02	.022
66 Uruguay	-.55	.08	-6.88	-.39	-.71	.000	.06	.01	6.00	.08	.04	.000	-.31	.03	-1.33	-.25	-.37	.000	.26	.07	3.71	.40	.12	.000
67 USA	-.73	.09	-8.11	-.55	-.91	.000	.07	.02	3.50	.11	.03	.000	-.42	.04	-1.50	-.34	-.50	.000	.1	.06	1.67	.22	-.02	.096
68 Vietnam	-.25	.07	-3.57	-.11	-.39	.000	.05	.02	2.50	.09	.01	.012	-.15	.03	-5.00	-.09	-.21	.000	.24	.05	4.80	.34	.14	.000

Note. The effect sizes are for four big-fish-little-pond effect studies based on PISA data collections between 2000 and 2012, which are reported in previously published results: PISA200 (Marsh & Hau, 2003: 103,558 students from 26 countries); PISA2003 (Seaton, Marsh & Craven, 2009, 2010: 265,180 students from 41 countries); PISA2006 (Nagengast & Marsh, 2012: 397,500 students from 57 countries); and PISA2012 (Marsh, Parker & Pekrun, 2017: 485,490 from 68 countries/regions), including the PISA2012 data considered in the present investigation. Coefficients in red are not statistically significant

Supplemental Table 4.

Parameter Estimates and Standard Errors (SE) for Each of 68 Countries/Regions (An Extension of Table 3 in the Main Text)

	Intercept		L1 Math Achieve (MA)		L2 Math Achieve		L1MA by L2MA		Relative School Year		L1-Read Achieve		L1-Science Achieve		L2-Read Achieve		L2-Science Achieve	
Country/Region	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
1 Albania	-.19	.06	.11	.05	-.28	.08	-.21	.05	.03	.02	.04	.02	-.07	.03	.04	.02	.06	.06
2 United Arab Emirates	.24	.02	.76	.04	-.68	.05	-.17	.02	.04	.01	-.27	.03	-.04	.03	-.27	.03	-.01	.05
3 Argentina	-.34	.04	.80	.05	-.64	.08	-.15	.04	.07	.01	-.20	.03	-.10	.03	-.20	.03	.14	.06
4 Australia	.06	.02	1.13	.03	-.64	.06	-.06	.01	.08	.01	-.51	.03	-.16	.03	-.51	.03	.05	.05
5 Austria	-.06	.03	1.09	.05	-.44	.08	-.03	.03	.07	.02	-.25	.04	-.28	.04	-.25	.04	.06	.06
6 Belgium	-.08	.02	.92	.04	-.53	.06	-.06	.02	.07	.01	-.42	.03	-.12	.03	-.42	.03	.02	.06
7 Bulgaria	-.19	.02	.68	.04	-.32	.07	-.04	.02	.02	.02	-.20	.03	-.14	.03	-.20	.03	.07	.06
8 Brazil	-.36	.03	.79	.05	-.51	.05	-.15	.04	.02	.01	-.29	.02	-.13	.02	-.29	.02	-.01	.04
9 Canada	.20	.02	1.11	.03	-.71	.04	.01	.03	.10	.01	-.29	.02	-.23	.02	-.29	.02	.15	.04
10 Switzerland	.06	.02	1.13	.03	-.65	.06	-.01	.03	.09	.01	-.63	.02	-.17	.03	-.63	.02	.08	.05
11 Chile	-.12	.02	1.16	.04	-.57	.08	-.10	.03	.09	.01	-.23	.03	-.29	.03	-.23	.03	.09	.06
12 Colombia	-.01	.03	.71	.07	-.45	.06	-.18	.05	.05	.01	-.17	.03	-.12	.03	-.17	.03	.00	.05
13 Costa Rica	.13	.04	.77	.06	-.36	.08	-.17	.06	.05	.01	-.25	.03	-.11	.03	-.25	.03	-.04	.06
14 Czech Republic	-.09	.03	1.06	.04	-.51	.06	-.03	.02	.07	.02	-.34	.03	-.19	.03	-.34	.03	.09	.06
15 Germany	.18	.03	1.26	.04	-.82	.07	-.03	.02	.07	.01	-.52	.04	-.19	.04	-.52	.04	.08	.06
16 Denmark	.14	.02	1.42	.04	-.78	.08	.05	.03	.07	.02	-.68	.03	-.14	.03	-.68	.03	-.02	.05
17 Spain	-.14	.02	1.00	.02	-.65	.04	.03	.02	.08	.01	-.29	.02	-.25	.02	-.29	.02	.10	.04
18 Estonia	.05	.04	1.01	.05	-.56	.09	.05	.05	.08	.02	-.31	.04	-.21	.04	-.31	.04	.07	.06
19 Finland	-.06	.02	1.16	.04	-.68	.07	.19	.04	.18	.01	-.46	.02	-.17	.03	-.46	.02	.02	.06
20 France	-.26	.03	1.14	.04	-.52	.07	.07	.02	.04	.02	-.46	.03	-.12	.04	-.46	.03	.00	.06
21 United Kingdom	.12	.02	1.27	.03	-.67	.07	-.03	.02	.07	.01	-.69	.03	-.14	.03	-.69	.03	.05	.05
22 Greece	.09	.03	.83	.04	-.39	.08	-.13	.03	.05	.02	-.25	.03	-.08	.03	-.25	.03	.04	.06
23 Hong Kong-China	-.14	.04	1.03	.05	-.65	.08	-.05	.04	.06	.01	-.58	.04	-.14	.04	-.58	.04	.10	.07
24 Croatia	-.28	.03	.98	.04	-.41	.08	-.02	.03	.03	.02	-.31	.04	-.29	.04	-.31	.04	.09	.07
25 Hungary	-.10	.03	1.10	.05	-.54	.08	.04	.02	.06	.02	-.33	.04	-.26	.04	-.33	.04	.05	.06
26 Indonesia	-.21	.04	.28	.06	-.27	.07	-.15	.04	.04	.01	-.06	.03	-.15	.04	-.06	.03	.04	.06

27 Ireland	.01	.03	1.15	.04	-.60	.08	.07	.04	.06	.01	-.35	.04	-.34	.04	-.35	.04	.16	.06
28 Iceland	.10	.03	1.15	.05	-.54	.10	.08	.05	.06	.03	-.39	.03	-.20	.04	-.39	.03	.00	.07
29 Israel	.33	.03	.77	.04	-.54	.08	-.14	.02	.05	.02	-.27	.03	-.14	.04	-.27	.03	.03	.06
30 Italy	.00	.01	.84	.02	-.22	.04	-.05	.01	-.01	.01	-.19	.02	-.18	.02	-.19	.02	-.12	.04
31 Jordan	.27	.04	.65	.05	-.32	.07	-.09	.03	.05	.02	-.09	.03	.02	.03	-.09	.03	.01	.06
32 Japan	-.45	.03	1.01	.04	-.52	.07	-.05	.02	.06	.03	-.42	.03	-.22	.03	-.42	.03	.08	.06
33 Kazakhstan	.19	.03	.42	.04	-.27	.06	-.19	.05	.03	.01	.02	.03	-.03	.03	.02	.03	-.05	.06
34 Korea	-.49	.03	.98	.05	-.48	.08	.00	.03	.06	.03	-.55	.04	-.15	.04	-.55	.04	.07	.07
35 Liechtenstein	.06	.08	1.03	.12	-.54	.13	-.03	.07	.07	.03	-.47	.10	-.13	.08	-.47	.10	.05	.08
36 Lithuania	-.08	.02	1.03	.04	-.51	.08	.04	.03	.10	.02	-.36	.04	-.14	.04	-.36	.04	.06	.06
37 Luxembourg	.01	.03	.96	.04	-.45	.11	.00	.03	.07	.01	-.46	.03	-.07	.03	-.46	.03	.07	.07
38 Latvia	-.10	.02	1.06	.04	-.51	.08	.02	.04	.06	.02	-.27	.04	-.32	.04	-.27	.04	.08	.06
39 Macao-China	-.27	.05	.86	.04	-.52	.09	-.01	.04	.08	.01	-.38	.03	-.19	.04	-.38	.03	.06	.07
40 Mexico	-.05	.02	.82	.04	-.43	.04	-.13	.05	.04	.01	-.21	.02	-.13	.02	-.21	.02	.03	.04
41 Montenegro	-.22	.05	.82	.07	-.44	.11	-.19	.09	.03	.02	-.28	.03	-.19	.04	-.28	.03	.06	.07
42 Malaysia	.01	.04	.74	.05	-.65	.07	-.15	.04	.07	.03	-.17	.03	-.25	.04	-.17	.03	.14	.07
43 Netherlands	.15	.03	1.14	.05	-.58	.08	-.07	.03	.11	.02	-.48	.04	-.23	.04	-.48	.04	.10	.06
44 Norway	-.17	.03	1.27	.04	-.61	.09	.11	.05	.07	.03	-.29	.03	-.25	.03	-.29	.03	.05	.06
45 New Zealand	.05	.03	.99	.04	-.59	.08	.00	.03	.09	.02	-.46	.03	-.15	.04	-.46	.03	.07	.06
46 Peru	-.24	.05	.72	.07	-.60	.07	-.16	.05	.05	.01	-.21	.03	-.14	.03	-.21	.03	.08	.06
47 Poland	-.05	.03	1.19	.04	-.72	.08	.08	.04	.11	.02	-.45	.03	-.25	.04	-.45	.03	.10	.07
48 Portugal	-.18	.02	1.00	.04	-.44	.08	-.02	.03	.04	.01	-.30	.03	-.25	.03	-.30	.03	.10	.06
49 Qatar	.01	.04	.51	.05	-.47	.07	-.23	.03	.03	.01	-.13	.02	-.06	.03	-.13	.02	-.01	.06
50 Shanghai-China	-.02	.06	.80	.05	-.44	.09	-.09	.03	.11	.01	-.65	.04	.04	.04	-.65	.04	.00	.07
51 Perm (Russian Fed)	-.03	.03	.66	.05	-.35	.10	-.01	.04	.06	.02	-.11	.04	-.18	.05	-.11	.04	.05	.07
52 Florida (USA)	.19	.04	1.10	.06	-.65	.11	-.01	.06	.06	.02	-.37	.05	-.24	.05	-.37	.05	.06	.07
53 Connecticut (USA)	.42	.04	.82	.06	-.41	.11	-.06	.05	.05	.02	-.31	.05	-.11	.05	-.31	.05	.04	.07
54 Massachusetts (USA)	.41	.04	1.04	.06	-.51	.11	-.05	.04	.08	.02	-.31	.05	-.34	.05	-.31	.05	.02	.07
55 Romania	-.06	.03	.43	.04	-.15	.07	-.05	.03	.05	.02	-.12	.03	-.08	.03	-.12	.03	.05	.06
56 Russian Federation	.05	.02	.76	.04	-.36	.07	-.05	.03	.06	.01	-.13	.03	-.23	.03	-.13	.03	.02	.06
57 Singapore	.15	.05	.96	.05	-.36	.09	-.08	.03	.02	.02	-.38	.04	-.36	.04	-.38	.04	.12	.06
58 Serbia	-.23	.03	.80	.04	-.27	.08	-.03	.03	.06	.03	-.26	.03	-.12	.03	-.26	.03	.12	.06
59 Slovak Republic	-.34	.03	.90	.04	-.45	.07	.01	.02	.06	.01	-.32	.03	-.20	.04	-.32	.03	.12	.06
60 Slovenia	-.13	.02	1.07	.04	-.54	.07	.00	.02	.06	.02	-.30	.04	-.29	.04	-.30	.04	.10	.06
61 Sweden	.07	.02	1.11	.05	-.61	.09	.05	.04	.10	.03	-.36	.03	-.17	.04	-.36	.03	.02	.06

62 Chinese Taipei	-.62	.04	1.14	.04	-.41	.07	-.07	.02	.06	.02	-.48	.04	-.45	.04	-.48	.04	.11	.07
63 Thailand	-.18	.02	.64	.04	-.42	.06	-.03	.02	.06	.01	-.29	.03	-.16	.03	-.29	.03	.06	.06
64 Tunisia	.07	.04	.65	.05	-.27	.08	-.17	.06	.07	.02	-.20	.03	-.07	.03	-.20	.03	.03	.06
65 Turkey	-.06	.03	.69	.04	-.29	.07	-.07	.03	.04	.01	-.10	.04	-.21	.04	-.10	.04	.04	.07
66 Uruguay	-.13	.03	.92	.05	-.55	.08	-.15	.05	.06	.01	-.31	.03	-.13	.03	-.31	.03	.08	.06
67 United States of America	.29	.03	1.27	.05	-.73	.09	-.09	.03	.07	.02	-.42	.04	-.33	.04	-.42	.04	.12	.06
68 Vietnam	-.13	.03	.49	.04	-.25	.07	-.02	.03	.05	.02	-.15	.03	-.18	.04	-.15	.03	.10	.06

Note. Results based on the random effects in Model 3 (Table 2) are presented separately for each of the 68 countries/regions. L1MA = Individual Student math achievement; L2MA = school-average math achievement (the big-fish-little-pond effect). Years = relative year-in-school effect. L1L2Int = Interaction between L1-Math achievement and L2-Math Achievement. Parameter estimates that differ from zero by more than two standard errors are statistically significant. Estimates are standardized effects in which all independent and dependent variables are standardized. All parameter estimates are significant at $p < .05$ when they differ from zero by more than 1.96 standard errors (SEs).

Supplemental Section 6.
Assumptions of Causality and Underlying Processes.

As mentioned in the main article, potential limitation of the present investigation is that it is based on a single wave of cross-sectional data; thus, causality cannot be inferred. Psychological, cross-cultural, and educational studies routinely must rely on such data, as it would be problematic and also unethical to randomly assign students to different schools, let alone different countries. Although it is appropriate to hypothesize causal relations, researchers need to interrogate support for causal hypotheses in relation to the construct validity of their interpretations (see Marsh, 2007) based on multiple methods, experimental designs, and time points, and testing their generalizability across diverse measures, settings, and countries and measures. However, particularly in relation to the BFLPE and the I/E models, this limitation has been explored in relation to several alternative research designs that provide a stronger basis for evaluating the causality implicit in the theoretical predictions.

BFLPE

A growing trend in BFLPE research has been the use of various combinations of longitudinal, quasi-experimental, and true experimental designs: all of these have been shown to support the BFLPE (see Marsh, 2007; Marsh & Seaton, 2015; Marsh, Seaton, et al., 2008; Nagengast & Marsh, 2012). Quasi-experimental, longitudinal studies based on matching designs as well as statistical controls, show that ASC declines when students move from mixed-ability schools to academically selective schools. This is based on pre-post comparisons and on comparisons with students matched on academic ability who continue to attend mixed-ability schools. There is also support for the BFLPE in secondary-school studies where achievement measures were administered before students began their secondary schooling (e.g., Marsh, Kong, Hau, 2000; Marsh, Parker & Pekrun, 2018). Extended longitudinal studies (Marsh et al., 2000; Marsh, Trautwein, Lütdke, Baumert, & Köller, 2007) show that the BFLPE becomes increasingly negative the longer students attend a selective school, and is maintained even 2 and 4 years after graduation from high school.

There is also good support for the convergent and discriminant validity of the BFLPE, in that it is largely limited to academic components of self-concept and is nearly unrelated to non-academic components of self-concept and to global self-esteem (Marsh, 2007; Marsh & Parker, 1984). As shown here, the BFLPE has good cross-cultural generalizability. Whilst the “third variable” issue is a potential threat to contextual studies without random assignment, Marsh and colleagues (Marsh & Hau, 2004; Marsh & Seaton, 2015; Marsh, Seaton et al., 2008) argue that this is an unlikely counter-explanation of BFLPE results, in that most potential “third variables” (resources, per student expenditures, SES, teacher qualifications, etc.) are positively related to school-average achievement, so that controlling for them would increase the size of the BFLPE. Finally, Zell and Alicke (2009; see also Alicke, Zell, & Bloom, 2010) have provided support for the processes underlying the BFLPE by experimentally manipulating the frame-of-reference through giving feedback to participants about their performances compared to others. Consistent with BFLPE field studies, in a variety of different conditions they reported results from a variety of experimental manipulations in which higher frames of reference resulted in lower self-evaluations.

Internal/External Frame-of-Reference Model

In I/E research there is also additional evidence in support of the causal interpretations based on the model. Experimentally, Möller and Köller (2001) and Pohlmann and Möller (2009) have shown how manipulated feedback on either verbal or math achievement in one subject area has an inverse effect on ASC in the other domain. The I/E model has also been validated using introspective diary studies and experimental designs. For example, in two diary studies, Möller and Husemann (2006) confirmed that students spontaneously carry out dimensional comparisons in everyday life, with contrast effects from one domain to self-evaluations and emotions in the other. Möller, Pohlmann, Köller, and Marsh (2009) further noted the complementary features of a comprehensive meta-analysis of I/E studies juxtaposed with a large cross-cultural PISA study. In support of the generalizability of the I/E predictions, results based on the PISA2000 data (the only PISA data collection to include both math and verbal psychological constructs) were highly consistent with the results based on a comprehensive meta-analysis of I/E studies.

Red Shirting

Parker, Marsh, Thoemmes, and Biddle (2018) specifically addressed the issue of causality in relation to the negative red shirting, using an instrumental-variable approach based on longitudinal data from the Longitudinal Study of Australian Youth. Importantly, consistent with observations in relation to other frame-of-reference effects, they found that the added control provided by the instrumental variable approach actually increased rather than decreased the size of positive red shirtings, and that support for the model extended to changes in self-beliefs over time. Indeed, they found that red shirtings positively predicted tertiary entrance ranks, which are the primary basis of university entrance and (actual) subsequent, university attendance. Consistent with the theoretical basis of the red shirting presented here, the positive effects of red shirtings were almost completely mediated by ASC. These results provide strong support for causal interpretations of the red shirting. Marsh (2016) provided evidence based on PISA data, that the positive red shirting can explain negative effects associated with starting school early or being accelerated, as well as the positive effects of starting school at an older age or repeating a year in school.

Expanding the implications of the red shirting, Marsh, Pekrun et al. (2017) found that repeating a year in school had largely positive effects for a diverse range of 10 outcomes (e.g., MSC, self-efficacy, anxiety, relations with teachers, parents and peers, school grades, and standardized achievement test scores). Indeed, the largest and most enduring positive effects of repeating a year in school were for school achievement in relation to school grades. Importantly, gains in the first year following retention were retained in subsequent school years. Although the authors emphasized frame-of-reference effects and social comparison theory, they also argued that the results are consistent with implications arising from mastery learning research and the Mathew Effect. Mastery learning research suggests that giving students sufficient time to master new materials will result in improved performances for slower learners, whereas the Mathew effect suggests that small differences at any stage of development will grow to become larger over time if there is no remedial intervention. In summary, the Marsh, Pekrun, et al. results provide good support for this interpretation of the potentially controversial positive red shirting, in respect to repeating a year in school. Particularly considering that the Marsh, Pekrun et al. results are contrary to at least some accepted wisdom about school retention, they have important implications for educational researchers, policy makers and parents.

Integration of Social and Dimensional Comparison Theory: BFLPE-Compensatory Effect

An important contribution of the present investigation was to bring together social comparison theory and dimensional comparison theory into an integrated meta-theory, providing a broader framework from which to study frame-of-reference effects more generally. In previous research, this sort of integration of the BFLPE and I/E model was first proposed by Marsh (1994; also see Marsh, 2007), who entertained the possibility of separate school-level contexts in relation to mathematics and verbal achievement. This early study provided preliminary support for the theoretical predictions, although appropriate multilevel perspectives to test the model more fully, were not readily available at that time. A similar combination of social comparison (as in the BFLPE) and dimensional comparison effects (as in the I/E model) was also evident in the laboratory studies of Pohlmann and Möller (2009).

Supplemental Section 7.
MLWIN Macro used to fit Model 1 in Table 2.

```
Note: start with a clean model and read in standard dataset
WIPE

LOAD "D:\Dropbox\herb\pisa\pisa2012\PISA2012_R\pisa2012 4NOV2016 MultImpv83+19L3 12May2017.wsz"
clear
note   pisa2012 4NOV2016 MultImpv83+19L3 12May2017.wsz
Note delete all previous stored models
MWIPE

MARK 0
BATCH 1
weight  1 2  c8
Note: Set up basic model with dep variable and level indicators
RESP  c18
IDEN 1 c2
IDEN 3 c6
IDEN 2 c7
ESTM 1
EXPA 2
CENT 0
Note: Add variables to be considered
ADDT  'cons'
CENT 0
SETV 3  c75
SETV 2  c75
SETV 1  c75
CENT 0
ADDT  'ZPVxmath'
CENT 0
CENT 0
ADDT  'L1AchSq'
CENT 0
CENT 0
ADDT  'L2Mach'
CENT 0
CENT 0
ADDT  'ZGRADE'
CENT 0
MACO 0
ADDT  'L1L2AchInt'
CENT 0
MACO 0
NoteADDT  'L2a_Grd'
Note CENT 0
Note MACO 0

ADDT  'L3MAch'
CENT 0
MACO 0
```

```
Addt 'ZPVxRead'
Addt 'ZPVxSCI'


WSET
MLAVerage  'nSCHLID'   'ZPVxSCI'   'c122'
NAME C122 "L2SciAch"
MLAVerage  'nSCHLID'   'ZPVxRead'   'c123'
NAME C123 "L2RAch"
ERASe C9997
NAME C9997 "c9997"
ERASe C9994
NAME C9994 "c9994"
ERASe C9993
NAME C9993 "c9993"
WSET

addt "L2SciAch"
addt "L2RAch"

Note: Add weights
NFMT 1 4
WSET
weight  1 2  c9997
weight  2 2  c9996
weight  3 2  c9995
Note: weights level 1 with standardized wts(1) in "c8"
weight  1 2  c8
PREF 0
POST 0
offsets 1
offsets 2
offsets 3
erase  c1091   c1090
mark 1  c1091
mark 1  c1090

Note: Use EXCL to select cases to be used
EXCL 1  c90

Note: Sets Random Structure-here random var & diagonal CV at L3
Note SETV 3   'ZPVxmath'


note becauseSETV 3   'L1AchSq'
note becauseCLRE 3  'cons'    'L1AchSq'
note becauseCLRE 3  'L1AchSq'  'ZPVxmath'

Note SETV 3   'L2Mach'


Note SETV 3    'ZGRADE'

Note SETV 3    'L1L2AchInt'

Note: make residual matrix diagonal
smat 2 0
```

```
smat 3 0


START
BATCH 1
like b100
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
ESTM 2
EXPA 2
MSTO '\model Xx1-5'
SEPIck C200
NAME c200 '11-6se(res)'
NOTE: ROUT 2 c201 c202
JOIN C1098 C1096 C210
NAME c210 'I1-6est(fi+rn)'
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
join c190 b100 c190
name c190 'likI1-5'

EXCL 1  c84
START
BATCH 1
PREF 0
POST 0
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join c190 b100 c190
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
MSTO '\model XXIMP0'
SEPIck C201
NAME c201 'I0 se(res)'
JOIN C1098 C1096 C211
NAME c211 'I0-est(fi+rn)'
join c191 b100 c191
name c191 'likI0'

EXCL 1  c85
START
BATCH 1
PREF 0
```

```
POST 0
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
MSTO '\model XXIMP1'
SEPIck C202
NAME c202 'I1se(res)'
JOIN C1098 C1096 C212
NAME c212 'I1-est(fi+rn)'
join c192 b100 c192
name c192 'likI1'

EXCL 1  c86
START
BATCH 1
PREF 0
POST 0
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
MSTO '\model XXIMP2'
SEPIck C203
NAME c203 'I2se(res)'
JOIN C1098 C1096 C213
NAME c213 'I2-est(fi+rn)'
join c193 b100 c193
name c193 'likI2'

EXCL 1  c87
START
BATCH 1
PREF 0
POST 0
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
```

```
MSTO '\model XXIMP3'
SEPIck C204
NAME c204 'i3se(res)'
JOIN C1098 C1096 C214
NAME c214 'I3-est(fi+rn)'
join c194 b100 c194
name c194 'likI3'

EXCL 1  c88
START
BATCH 1
PREF 0
POST 0
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
MSTO '\model XXIMP4'
SEPIck C205
NAME c205 'i4se(res)'
JOIN C1098 C1096 C215
NAME c215 'I4-est(fi+rn)'
join c195 b100 c195
name c195 'likI4'

EXCL 1  c89
START
BATCH 1
PREF 0
POST 0
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
moni 0
like b100
join  c1091 b100  c1091
join  c1090  c1098 c1096  c1090
moni 1
MSTO '\model XXIMP5'
SEPIck C206
NAME c206 'I5se(res)'
JOIN C1098 C1096 C216
NAME c216 'I5est(fi+rn)'
join c196 b100 c196
name c196 'likI5'

Note: apply Rubin Rules over imputations
CALCulate  'c217' = ( 'I1-est(fi+rn)' + 'I2-est(fi+rn)' + 'I3-est(fi+rn)' + 'I4-est(fi+rn)' +  'I5est(fi+rn)'  )/5
CALCulate 'c218' = ( 'I1se(res)'  +  'I2se(res)'  +  'i3se(res)'  +  'i4se(res)' +  'I5se(res)' ) /5
```

```
CALCulate 'c219' =( (  (('I1-est(fi+rn)' -  'c217')^2) +  (('I2-est(fi+rn)' -  'c217')^2) +  (('I3-est(fi+rn)' -  'c217')^2) +  (('I4-
        est(fi+rn)' -  'c217')^2) + (('I5est(fi+rn)'  -  'c217')^2) )/5 ) + 'c218'
CALCulate  'c220' = ( 'likI1' + 'likI2' + 'likI3' +'likI4' +'likI5')/5
NAME c217 'RubinEST'
NAME c218 'MnSE'
NAME c219 'RubinSE'
NAME c220 'MnLikHR'
```

$$\mathrm{ZSCMAT}_{ijk} \sim \mathrm{N}(XB, \Omega)$$

$$\mathrm{ZSCMAT}_{ijk} = \beta_{0ijk}\mathrm{cons} + \beta_1\mathrm{ZPVxmath}_{ijk} + \beta_2\mathrm{L1AchSq}_{ijk} + \beta_3\mathrm{L2Mach}_{jk} + \beta_4\mathrm{ZGRADE}_{ijk} + \beta_5\mathrm{L1L2AchInt}_{ijk} + \beta_6\mathrm{L3Mach}_{k} + \beta_7\mathrm{ZPVxREAD}_{ijk} + \beta_8\mathrm{ZPVxSCI}_{ijk} + \beta_9\mathrm{L2SciAch}_{jk} + \beta_{10}\mathrm{L2RAch}_{jk}$$

$$\beta_{0ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\left[v_{0k}\right] \sim \mathrm{N}(0,\ \Omega_v)\ :\ \Omega_v = \left[\sigma^2_{v0}\right]$$

$$\left[u_{0jk}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[\sigma^2_{u0}\right]$$

$$\left[e_{0ijk}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[\sigma^2_{e0}\right]$$

*-2*loglikelihood(IGLS Deviance)* = 1247475.1478(485490 of 2912940 cases in use)

Supplemental References

Note: References listed here are for articles cited in Supplemental Materials that are not also cited in the main text.

Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology, 54*, 403–425. doi:10.1146/annurev.psych.54.101601.145056

Gladwell, M. (2008), Outliers, New York: Little, Brown and Company.

Mislevy, R.J., Beaton, A., Kaplan, B.A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*, 133–161. doi:10.1111/j.1745-3984.1992.tb00371.x

Möller, J., & Husemann, N. (2006). Internal comparisons in everyday life. *Journal of Educational Psychology*, *98*, 342–353. doi:10.1037/0022-0663.98.2.342

Spitzer, S.; Cupp, R.; Parke, R.D. (1995). School entrance age, social acceptance, and self-perception in kindergarten and 1st grade. *Early Childhood Research Quarterly*. *10* (4), 433–450. doi:10.1016/0885-2006(95)90015-2.

Von Davier, M., Gonzalez, E., & Mislevy, R.J. (2009). What are plausible values and why are they useful? IERI Monograph Series. *Issues and Methodologies in Large-Scale Assessments*, *2,* 9–36.