Journal of Educational Psychology

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

Benjamin Nagengast and Herbert W. Marsh

Online First Publication, April 9, 2012. doi: 10.1037/a0027697

Journal of Educational Psychology

0022-0663/12/$12.00 DOI: 10.1037/a0027697

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

Benjamin Nagengast
University of Tübingen and University of Oxford

Herbert W. Marsh
University of Oxford, University of Western Sydney, and King Saud University

Being schooled with other high-achieving peers has a detrimental influence on students' self-perceptions: School-average and class-average achievement have a negative effect on academic self-concept and career aspirations—the big-fish-little-pond effect. Individual achievement, on the other hand, predicts academic self-concept and career aspirations positively. Research from Western and developed countries implies that the negative contextual effect on career aspirations is mediated by academic self-concept. Using data from the Program for International Student Assessment (PISA) 2006 (a total of 398,750 15-year-old students from 57 countries), we test the generalizability of this mediation model in science using a general multilevel structural equation modeling framework. Individual achievement was positively related to academic self-concept (52 countries) and career aspirations (42 countries). The positive effect on career aspirations was mediated by self-concept in 54 countries. The negative effects of school-average achievement on self-concept (50 countries) and career aspirations (31 countries) also generalized well. After controlling for self-concept at both the individual and the school level, there were significant indirect contextual effects in 34 countries—evidence for mediation of the contextual effect of school-average achievement on career intentions by academic self-concept.

Keywords: academic self-concept, career aspirations, big-fish-little-pond effect, mediation, cross-cultural studies

Supplemental materials: http://dx.doi.org/10.1037/a0027697.supp

Cultivating career aspirations is an important educational objective. Good schooling and teaching should not only further the acquisition of knowledge in the compulsory secondary school setting but also inspire students to pursue more specialized knowledge at the postsecondary level. Fostering career intentions in the sciences is particularly important, because industrialized countries currently suffer from a shortage of well-trained scientists and other professionals in related fields like engineering (Organisation for Economic Co-operation and Development [OECD], 2007).

Research consistently emphasizes the importance of individual school achievement and self-beliefs as determinants of career choice (e.g., Betz & Hackett, 1981; Hackett & Betz, 1981; Marsh, 1991; Marsh & Yeung, 1997a). Academic self-concept, the perception of a person's abilities in a specific subject (Byrne & Shavelson, 1986), has been shown to be the most important predictor for course and career choices—even more important than individual achievement when both are controlled simultaneously (Guay, Larose, & Boivin, 2004; Marsh & Yeung, 1997a). However, it is not only individual factors that shape students' career intentions and choices. A well-established research tradition has also focused on the contribution of environmental and contextual factors. In an early review of this field, Alwin and Otto (1977) reported that school-average achievement influences career intentions negatively. Similar findings on students' academic self-concept have come from big-fish-little-pond effect (BFLPE) research (for a comprehensive review, see Marsh, Seaton, et al., 2008). BFLPE research demonstrates that school-average achievement has a negative effect on academic self-concept after controlling for individual achievement. In mediational analyses (Marsh, 1991; Marsh & O'Mara, 2010), the negative effect of school-average achievement on career aspirations is substantially reduced—or disappears altogether—when academic self-concept is controlled.

Benjamin Nagengast, Center for Educational Science and Psychology, Department of Education, University of Tübingen, Tübingen, Germany; Department of Education, University of Oxford, Oxford, United Kingdom. Herbert W. Marsh, Department of Education, University of Oxford; Centre for Positive Psychology and Education, University of Western Sydney, Sydney, Australia; Department of Education, King Saud University, Riyadh, Saudi Arabia.

Parts of this research were presented at the 14th Biennial EARLI Conference for Research on Learning and Instruction 2011, Exeter, United Kingdom. This research was supported by a professorial fellowship from the Economic and Social Research Council to Herbert W. Marsh. We thank Geert Hofstede for making the cultural orientation scores available to us. We also thank Kou Murayama and Reinhard Pekrun for comments on previous drafts of this article.

Correspondence concerning this article should be addressed to Benjamin Nagengast, Center for Educational Science and Psychology, Department of Education, University of Tübingen, Europastrasse 6, 72072 Tübingen, Germany. E-mail: benjamin.nagengast@uni-tuebingen.de

In this article, we extend existing research in three important ways:

1. Research on the relations between academic achievement, academic self-concept, and career intentions has been mostly confined to a few individualistic and developed countries and has not taken a cross-cultural comparative perspective. We investigate these relations in representative samples of 15-year-old students from 56 diverse countries—a larger and more diverse sample than in any other previous BFLPE study. Thus, we provide important insights into the ubiquitous nature of frame-of-reference effects for outcomes other than academic self-concept and the mechanisms that bring them about.

2. We test the relation between achievement, academic self-concept, and career intentions in the subject area of science. While science has been an important area of study with respect to the effects of individual achievement on career aspirations (see Fouad, 2007), there have been few BFLPE studies in science, and apparently there are none that have incorporated aspirations in this subject area with important economic and policy implications.

3. Recent developments of multilevel structural equation models (SEMs; Lüdtke et al., 2008; Marsh, Lüdtke, et al., 2009; Mehta & Neale, 2005) allow researchers to simultaneously control for measurement error and sampling error and have led to the development of a general framework for multilevel mediation analysis (Preacher, Zhang, & Zyphur, 2011; Preacher, Zyphur, & Zhang, 2010). We use this framework to address some of the methodological shortcomings of earlier studies (failure to account for the multilevel structure, Marsh, 1991; failure to control for measurement error, Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006) to create a substantive–methodological synergy (Marsh & Hau, 2007).

The Substantive Focus: Relations Between Achievement, Academic Self-Concept, and Career Aspirations

Academic Self-Concept

The beneficial effects of self-concept—a person's sense of self, shaped through interaction with the environment and other people (Shavelson, Hubner, & Stanton, 1976)—are well-documented across many areas of psychology (see Marsh, 2006; Marsh & Craven, 2006, for a review). In the educational domain, a positive academic self-concept (ASC), defined as the self-perception of one's academic abilities and competencies (Byrne & Shavelson, 1986) has—among others—been linked to coursework selection, engagement, intrinsic motivation, subsequent achievement, educational aspirations, and subsequent university attendance (e.g., Guay, Larose, & Boivin, 2004; Marsh, 1991; Marsh & Craven, 2006; Marsh & Yeung, 1997a, 1997b; also see meta-analysis by Valentine, DuBois, & Cooper, 2004).

Big-fish-little-pond effect. Although the effect of individual achievement (ACH) on ASC has been shown to be unequivocally and reciprocally positive (e.g., Marsh & Craven, 2006; Marsh & Yeung, 1997b; Valentine & DuBois, 2005; Valentine et al., 2004), it is well documented that selective school systems and ability grouping or streaming have detrimental consequences for the ASC of high-achieving students. Students in classes or schools with high average ACH have lower ASCs than their equally able peers in schools or classrooms with average or low ACH—the BFLPE (e.g., Craven, Marsh, & Print, 2000; Marsh, 1991; Marsh, Seaton, et al., 2008). Marsh and colleagues (e.g., Marsh, 1991; Marsh, Seaton, et al., 2008) have argued that the BFLPE is based upon a social comparison process (for a detailed account of the theoretical background of the BFLPE, see Marsh, Seaton, et al., 2008). Students form their ASC by comparing their own achievement to the achievement of their classmates, using the class average as frame of reference. This process tends to have detrimental consequences for high-achieving students in selective academic environments where they mix with other high-achieving students, leading to a more negative perception of their own ACH and more negative ASCs. In less selective environments, the abilities of the peers will have a higher variability and be lower on average, leading to a more positive self-evaluation and more positive ASCs. Students with lower achievement levels, on the other hand, can benefit from selective schooling because when they are grouped with other less achieving students, their frame of reference contains, on average, less able peers. Hence, their ASC will be more positive than if they had been placed in a classroom with high-achieving students.

Evidence from cross-cultural studies. Evidence for this negative contextual effect of school-average ACH on ASC (controlling for the positive effect of individual ACH)—the BFLPE—is strong and has been accumulating for more than two decades (for a comprehensive review, see Marsh, Seaton, et al., 2008). The pervasiveness of the BFLPE has been demonstrated across student groups, subject domains, and cultures. Most BFLPE research has been undertaken in Western developed countries (e.g., Craven et al., 2000; Marsh, 1987, 1991, 2004; Marsh, Chessor, Craven, & Roche, 1995; Marsh, Köller, & Baumert, 2001; Marsh & Parker, 1984; Mulkey, Catsambis, Steelman, & Crain, 2005; Nagengast & Marsh, 2011; Seaton, Marsh, et al., 2008; Trautwein, Lüdtke, Marsh, & Nagy, 2009; Tymms, 2001; Zeidner & Schleyer, 1999). Only few studies have particularly focused on Asian countries and school systems (e.g., Liem, Marsh, Martin, McInerney, & Yeung, 2010; Liu, Wang, & Parkins, 2005; Marsh, Kong, & Hau, 2000). Almost none of these studies provide any evidence for substantial moderating effects of a wide variety of additional school- and individual-level variables that have been considered by Marsh (2006; see also Seaton, Marsh, and Craven, 2010).

The studies by Marsh and Hau (2003) and Seaton, Marsh, and Craven (2009) stand out as the two most powerful cross-cultural comparative investigations of the BFLPE. These studies demonstrated that the BFLPE generalized across different countries and cultures with nationally representative samples of students and hence established its status as a "pan-human theory" (Seaton et al., 2009, p. 403). Using data from the Program for International Student Assessment (PISA) 2000 (OECD, 2001), focusing on general ASC, Marsh and Hau (2003) were the first to show the BFLPE in nationally representative samples of 15-year-old students from 26 countries. They demonstrated that the effect of school-average ACH on ASC was significantly negative in 24 out of the 26 countries participating in PISA 2000. Marsh and Hau also found a small but significant variation of the BFLPE across the countries in their sample. However, although they used the largest available sample at that time, the countries included in the Marsh and Hau study were predominantly developed and individualistic

countries—restricting the generalizability of their findings to a relatively narrow subset of cultures and stages of socioeconomic development.

Seaton et al. (2009) considerably extended the analyses by Marsh and Hau (2003). Building upon theoretical arguments and limited experimental evidence (McFarland & Buehler, 1995), they reasoned that the BFLPE was likely to be stronger in individualistic cultures (Hofstede, Hofstede, & Minkov, 2010). In collectivist countries, the social group to which one belongs is more important than the individual; thus, students may see being in an academically selective environment as an asset and assimilate their ASC to their context. In individualist countries, the individual takes precedence over the social group, and students are more likely to contrast their ASC away from the context of their peers. Hence, Seaton et al. (2009) predicted that students in individualistic societies would suffer a stronger BFLPE. Seaton et al. used data from the international sample of PISA 2003 (OECD, 2005) focusing on ASC in mathematics. Their sample included 41 countries with more variation in cultural orientation and economic development than in the study by Marsh and Hau (2003). Again, the BFLPE generalized across most of the countries included in the sample, being statistically significant in 38 of the 41 countries. As predicted, they also found support for a small interaction effect between country-level individualism and the BFLPE: The BFLPE was slightly larger for students from individualistic countries compared to students from collectivist countries. However, the effect size of this interaction was very small. Economic development, another country-level predictor, did not interact with school-average achievement, indicating that the BFLPE generalized over countries at different stages of economic development.

Three shortcomings of Seaton et al.'s (2009) analyses are addressed in the present study. First, although more comprehensive than Marsh and Hau (2003), their sample was still relatively limited with respect to the variability of the economic development. According to their classification on the basis of gross national income per capita, all but five (out of 41) countries were classified as high or upper middle income countries. Using more sophisticated measures of economic and social development and a still richer set of countries may further unveil support for economic predictors of the BFLPE.

Second, Seaton et al. (2009) only studied the effects of a single cultural orientation variable (individualism) and restricted their analysis to the BFLPE. Meta-analyses of the effects of cultural dimensions (Taras, Kirkman, & Steel, 2010) suggest that other cultural dimensions are as influential as individualism for a variety of important psychological variables, supporting calls for consideration of more cultural dimensions in cross-cultural studies (e.g., Gelfand, Erez, & Aycan, 2007; Kirkman, Lowe, & Gibson, 2006; Tsui, Nifadkar, & Ou, 2007). Along these lines, Chiu and Klassen (2010) recently tested the influence of a larger set of cultural predictors on the relation between individual ASC and ACH, using data from PISA 2003. They found support for the importance of economic development, egalitarianism, masculinity and uncertainty avoidance as moderators of the relation of the effect of ASC on ACH. Individualism did not emerge as a significant moderator, presumably because of its strong correlation with economic development. However, their study did not include school-average ACH. To further uncover potential cultural predictors of the BFLPE, we extend Seaton et al.'s (2009) analysis by exploring the influence of all the initial five of Hofstede et al.'s (2010) cultural dimensions (individualism, power distance, uncertainty avoidance, masculinity and long-term orientation) on both the strength of the relation between individual ACH and ASC and the strength of the BFLPE.

Finally, Seaton et al. (2009) restricted their analysis to ASC. While this is in line with most BFLPE research, frame-of-reference effects such as the BFLPE have been shown to be operating for other important academic outcomes as well (Alwin & Otto, 1977; Marsh, 1991; Marsh & O'Mara, 2010; Preckel, Götz, & Frenzel, 2010; Trautwein et al., 2006). However, in contrast to academic self-concept, there is a lack of cross-cultural comparative studies for these outcomes. In addition, although there is evidence for the mediating role of ASC in bringing about the negative contextual effects on other outcomes (see Marsh, Seaton, et al., 2008), the extent to which these effects generalize across cultures and stages of socioeconomic development is largely unknown. Thus, we now introduce the second substantive focus of this study: career aspirations.

Career Aspirations and Career Choice

Research on career aspirations for the future (FUT)—students' hopes and expectations of pursuing a career in a specific field—has been dominated by the search for personal factors that influence career and occupational choices and by identifying ways to best match a person's preferences and the work environment (see e.g., Fouad, 2007; Holland, 1997; Savickas, 2002, 2005; Spokane, Cruza-Guet, Brown, & Lent, 2005; Super, 1980). Academic self-beliefs have been found to be an important predictor for career decisions at the individual level (Betz & Hackett, 1981; Hackett & Betz, 1981; Marsh, 1991; Marsh & Yeung, 1997a). Betz and Hackett (1981), for example, showed that women's preference for traditionally female occupations could not be explained by differences in objectively measured qualifications and achievement but could be related to their low self-beliefs in relation to these occupations. The social–cognitive model of career choice (Lent, Brown, & Hackett, 1994) assumes a mediational chain in which self-related cognitions influence interests, which in turn influence actions and choices to engage in an area which, in turn, predicts levels of performance. Empirical tests have generally supported these hypothesized relations in diverse populations (Ferry, Fouad, & Smith, 2000; Fouad & Smith, 1996; Lent et al., 2001, 2005; Lopez, Lent, Brown, & Gore, 1997; see also Trautwein et al., 2006; Wigfield & Eccles, 2000).

While individual self-perceptions along with a variety of other individual factors have been studied extensively as predictors of career aspirations and career development (see Fouad, 2007, for a review) and contextual barriers have also been an important focus (e.g., Lent et al., 2005), school and peer effects have received less attention in recent research on career choice and career development (for notable exceptions, see Jonsson & Mood, 2008; Pustjens, van de Gaer, Van Damme & Onghena, 2004). This is in stark contrast to a sociological research tradition dating back to the 1960s that focused strongly on school-related determinants of career ambition, such as school-average ability, school-average socioeconomic status, or racial composition (e.g., Davis, 1966; see review in Alwin & Otto, 1977). Similar to findings in the BFLPE literature, research in this tradition has consistently emphasized a

negative contextual effect of school-average achievement on career aspirations. Students who attend schools with a high average ability level expressed lower career aspiration compared to similar students who attended schools with lower average ability levels. These findings prompted Alwin and Otto (1977) to call for specific tests of the processes and variables that mediate the relation of school contextual effects and occupational aspirations.

Their call was answered in part by Marsh (1991), who combined the sociological research tradition that focused on career aspirations with his research program on the BFLPE on ASC. Using data from the longitudinal High School and Beyond study (National Center for Educational Statistics, 1986), Marsh showed that school-average ACH negatively influenced career aspirations after controlling for individual ACH and other individual differences. More important, though, he demonstrated that the effects of individual and school-average ACH on career aspirations were almost completely mediated by ASC. When ASC was controlled, the direct effects of school-average and individual ACH on career aspirations were no longer statistically significant. In a similar analysis using data from the longitudinal Youth in Transition study (Bachman, 1975), Marsh and O'Mara (2010) found negative long-term effects of school-average ability on educational and occupational aspirations over a span of 8 years. Importantly, these effects were mediated almost completely by ASC. After controlling for ASC, the effects of school-average ACH on aspirations were essentially zero.

While these studies provided initial evidence for the important role of ASC in mediating the negative effects of school-average achievement on career intentions, there are certain methodological shortcomings. No mediation studies to date have disentangled mediation effects at different levels (e.g., school vs. student levels). Also, while most recent studies used multilevel modeling techniques to account for the nesting of students within schools, there has been no BFLPE mediation study that simultaneously controlled for measurement error and the multilevel structure at the same time. New developments in multilevel structural equation modeling and mediation analysis reviewed in the following section offer exciting opportunities to address these questions.

Methodological Focus: Doubly Latent Contextual Effect Models of Mediation Effects

Statistically, the BFLPE is a contextual effect: that is, the effect of school composition in ACH on ASC when the effects of individual ACH have been statistically controlled. School-average ACH, a contextual variable formed by aggregating individual ACH (L1) to the school-level (L2), predicts systematic differences in ASC that remain after individual ACH has been controlled (Lüdtke, Marsh, et al., 2008; Lüdtke, Robitzsch, Trautwein, & Kunter, 2009; Marsh et al., in press). Operationally, this is equivalent to modeling individual ACH and its L2 aggregate as predictor variables for ASC (Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995). Recent methodological advances in the analyses of contextual effects (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Lüdtke, Marsh, et al., 2008) warrant a fresh look at the BFLPE.

Doubly Latent Contextual Effect Models

Methodological developments and substantive progress go hand in hand; this is the essence of substantive–methodological synergies (Marsh & Hau, 2007). Research on the BFLPE has gradually moved from employing single-level models (e.g., Marsh, 1984, 1987, 1991; Marsh & Parker, 1984) to more appropriate multilevel modeling techniques (e.g., Marsh & Hau, 2003; Marsh & Rowe, 1996; Seaton et al., 2010). Most BFLPE research, however, has not used latent variable models that address the problem of measurement error (e.g., Bollen, 1989; Kaplan, 2000) in the manifest indicators of ASC and ACH (for a notable exception, see Marsh, 1994). Measurement error occurs when a latent unobserved construct can only be assessed with a number of fallible indicators. In multilevel models, measurement error can occur at all levels of the hierarchy—that is, at the student level (L1) and the classroom or school level (L2). Appropriately controlling for its effects is crucial in the analysis of contextual effects such as the BFLPE. The bias introduced by unaccounted measurement error at L1 can lead to biased contextual effects—even creating the illusion of a contextual effect when that effect is essentially zero in the population (Harker & Tymms, 2004; Marsh, Seaton, et al., 2010).

Lüdtke et al. (2008; also see Shin & Raudenbush, 2010) identified a second source of bias in multilevel models—sampling error—that is particularly relevant for the analysis of contextual effects and is not addressed by conventional multilevel modeling techniques (e.g., Hox, 2002; Raudenbush & Bryk, 2002). Sampling error in aggregated L2 variables occurs when not all students from a class are included in the sample. In this case, the observed classroom or school mean of the L1 variable will not be equal to the true classroom or school mean, leading to biased estimates of the contextual effect. Drawing upon BFLPE research and the methodological advancements in Lüdtke et al. (2008), Lüdtke et al. (2011) and Marsh, Lüdtke, et al. (2009) introduced a doubly latent multilevel SEM for contextual effects that controls measurement error at L1 and L2, as well as sampling error due to not sampling all L1 units from an L2 unit (see also Goldstein & McDonald, 1988; McDonald, 1993, 1994; Mehta & Neale, 2005; Rabe-Hesketh, Skrondal, & Pickles, 2004). The model consists of confirmatory factor analysis models at L1 and L2 controlling measurement error on both levels by using multiple indicators for the considered constructs. Contextual variables (i.e., L2 aggregates of individual-level variables) are formed by a latent aggregation procedure that takes into account that the observed values of these variables are not equal to true population values if sampling error is present. Marsh, Lüdtke, et al. (2009) showed with a didactic example that the size of the estimated BFLPE could change substantially compared to the standard multilevel regression model and several partial correction models.

Multilevel Mediation Models

In multilevel mediation analysis, there has been a distinction between mediational chains that involve L1 and L2 processes (Krull & MacKinnon, 2001; MacKinnon, 2008; Preacher et al., 2010). This distinction is particularly relevant for the present study, as previous research has shown that some of the positive effects of L1-ACH and the negative contextual effects of L2-ACH on FUT are transmitted by ASC (Marsh, 1991; Marsh & O'Mara,

2010). Building on the multilevel contextual models by Lüdtke et al. (2008, 2011), Preacher et al. (2010) proposed a unified multilevel structural equation modeling framework for multilevel mediation analysis. Preacher et al. argued that multilevel mediation researchers have overlooked that L1- and L2-effects of variables can differ (i.e., contextual effects can be present) and that appropriately modeling mediational processes in multilevel situations would require disentangling ("unconflating"; p. 224) the effects at each level of the hierarchy. Thus, a properly specified multilevel mediation model should distinguish between effects at both L1 and L2.

Preacher et al. (2010) only separated L1- and L2-effects (not controlled for individual differences) into direct and indirect effects. In the present study, however, we are particularly interested in mediation of the negative contextual effect of L2-ACH (that is controlled for individual differences). In case of the BFLPE, the hypothesized mediational relation can be classified as a cross-level 2-1/2-1 chain (Krull & MacKinnon, 2001; Pituch & Stapleton, 2011; but not included in the framework of Preacher et al., 2010,), meaning that L2-ACH (an L2 variable) influences ASC (a variable with components both at L1 and L2), which in turn affects FUT (an L1 variable). Hence, the negative contextual effect of ACH on FUT can be transmitted both by the L1 component of ASC and by the aggregated L2 component. The derivation and presentation of the mediated effects have to take these cross-level effects into account if the relative importance of the mediational pathways is to be disentangled appropriately (see Appendix A in the supplemental materials for further formal derivations of contextual indirect effects and how they can be computed from the parameters of a multilevel SEM).

The Present Investigation

In the current study, we set out to address three shortcomings of previous BFLPE studies. We study the relation of ACH, ASC, and FUT in large representative samples from 57 countries, focus on the subject area of science, and apply recent methodological advances in multilevel structural equation modeling.

More specifically, we consider the contextual effect of ACH on ASC and FUT, focusing on its cross-cultural generalizability and on the mediational relationship between these three constructs. We first demonstrate the negative contextual effects of L2-ACH on ASC in the largest and most diverse set of countries yet considered (see also Nagengast & Marsh, 2011). While the negative contextual effect of L2-ACH on ASC has already been shown to generalize very broadly over different cultures and stages of economic development (but not for science, the focus of the present study), less is known about the ubiquity of the negative contextual effects of L2-ACH on FUT. Hence, we test whether negative contextual effects of L2-ACH on FUT can be found in the countries participating in PISA 2006. While the generalizability of these effects would be of great interest in itself, we also focus on the mediating role of ASC in this process. Given that the research on the mediating effects of ASC seems to have been exclusively conducted in Western countries, this investigation is timely and adds to the importance of ASC as an important outcome in itself and a facilitator of other desirable educational outcomes. Finally, to the extent that the effects differ between the countries, we explore the relation of the country-specific effects to potential country-level moderators, socioeconomic developmental status, and five cultural dimensions (individualism, power distance, masculinity, long-term orientation and uncertainty avoidance).

Method

Participants

PISA 2006 (OECD, 2007) assessed the competencies of 15-year-old students in OECD and partner countries focusing on science ACH using nationally representative samples. In the total international sample, data of 398,750 15-year-old students from 14,365 schools in 57 countries were available (OECD, 2007). Due to the sample size requirements of the multigroup multilevel SEMs, data from Liechtenstein (where only 339 students going to 12 schools were assessed) had to be excluded from the cross-cultural analyses. Here, the number of schools was smaller than the number of unique elements in the L2-covariance matrix, resulting in nonconvergence of multigroup SEMs. This resulted in a final sample of 398,411 students from 30 OECD and 26 partner countries for the cross-country comparisons.

Individual-Level Measures

Science achievement. Science achievement (ACH) in PISA 2006 was assessed as part of a 120-min testing session. Students were randomly assigned to one of 13 test booklets constructed according to a balanced incomplete block design. The booklets contained a variable number of problems in science, reading, and mathematics, with approximately 54% of the average testing time devoted to science problems. The problems used open- and closed-format answering options (for details, see OECD, 2007, 2009a). The assessment of science ACH was targeted at assessing how well students were able to "identify scientific issues, explain phenomena scientifically [and] use scientific evidence" (OECD, 2006, p. 20). Items covered the content areas physical systems, living systems, earth and space systems, and technology systems (OECD, 2007). Science ACH at the individual student level is represented by five plausible values (Mislevy, 1991; Mislevy, Beaton, Kaplan, & Sheehan, 1992; von Davier, Gonzalez, & Mislevy, 2009) that represent the individual ability distribution of the student and facilitate secondary analysis. Correct analyses of plausible values require that all models are run separately for each plausible value and the results integrated using principles of multiple imputation analysis (see OECD, 2009a, 2009b; and Appendix A in the online supplemental materials for this article).

Academic self-concept in science. After the achievement test, students completed a background questionnaire (for details, see OECD, 2007, 2009a) that included six items (e.g., "Learning advanced school science topics would be easy for me") that measured academic self-concept in science. The items were scored on a 4-point Likert scale ranging from 1 (*strongly agree*) to 4 (*strongly disagree*). For the present analyses, all items were reverse scored, so that higher values represent more positive ASCs. The median reliability (Cronbach's α) was 0.92 in the OECD countries and 0.87 in the partner countries (OECD, 2009a).

Career aspirations in science. The background questionnaire also included four items that assessed FUT in science (e.g.,

"I would like to work in a career involving broad science"). These items were also scored on a 4-point Likert scale ranging from 1 (*strongly agree*) to 4 (*strongly disagree*). For the present analyses, all items were reverse scored, so that higher values represent higher FUT. The median reliability was $\alpha = .92$ in the OECD countries and $\alpha = .90$ in the partner countries (OECD, 2009a). Full details of the scale construction and initial validation are given in the PISA 2006 technical report (OECD, 2009a; also see Marsh, Hau, Artelt, Baumert, & Peschar, 2006).

Country-Level Predictors

Socioeconomic development. Socioeconomic development was represented with the U.N. Human Development Index (United Nations Development Program, 2009), which combines an index of the adult literacy rate, the school enrollment ratio, the gross domestic product per capita, and measures of life expectancy in one index. For the purpose of this study, we used values from 2006, the year of the data collection.

Cultural dimensions. Scores on the original five cultural dimensions (individualism–collectivism, power distance, uncertainty avoidance, masculinity, and long-term orientation) proposed by Hofstede et al. (2010) were used as a second set of country-level predictors. All values were obtained from Hofstede et al. (2010). Nine countries had at least one missing value on the cultural dimensions and were excluded only from the analyses that related country-level predictors to country-specific effects of ACH. Further description of the scales and example items are given in the supplemental materials (Appendix A).

Data Analysis

Missing data were handled using multiple imputation (see Graham, 2009; Schafer, 1997; see supplemental materials, Appendix A). The amount of missing data for questionnaire items was small (ranging from 2.3% to 2.4% per item for the four indicators of FUT and from 7.9% to 8.5% per item for the six indicators of ASC). The ACH data were complete. All analyses used the survey weights provided in the PISA 2006 database to account for the complex sampling design and obtain representative results for the national population of 15-year-old students. Variables were standardized and centered around their country mean prior to the analyses to put them on the same scale and remove variance at the country level. All analyses were conducted using the multilevel structural equation module of Mplus (Version 6.1; Muthén & Muthén, 2010). For the total group analyses, two-level models were implemented that accounted for the nesting of students within schools. Countries were treated as a stratification variable so that standard errors and test statistics were corrected for the third level of nesting. In the multicountry analysis, two-level models with invariance constraints (see below) were fitted within countries treating country as a fixed grouping variable. More details about the imputation procedure, weighting, standardization, and assessment of model fit are given in the supplemental materials (Appendix A). All significance tests of individual parameters were conducted at the 5% level.

Multilevel confirmatory factor analysis. Following recommendations by Ryu and West (2009), we first tested the hypothesized factor structure of ASC and FUT separately at L1 (students) and L2 (schools). These tests showed that the two-factor solution held both at L1 ($\chi^2_{PS\_W} = 2{,}252.3$, $df = 34$, $RMSEA_{PS\_W} = 0.013$) and at L2 ($\chi^2_{PS\_B} = 175.9$, $df = 34$, $RMSEA_{PS\_B} = 0.003$). In order to establish measurement invariance (Marsh, Muthén, et al., 2009; Mehta & Neale, 2005; Meredith, 1993; Widaman & Reise, 1997) of ASC and FUT between L1 and L2, a series of invariance tests was then conducted using a multilevel confirmatory factor analysis model. Cross-level measurement invariance ensures that the latent variables at L2 can be interpreted as the school-level means of the latent variables at L1. This invariance ensures that the variance decomposition into L1 and L2 variance is meaningful for the latent variables and that contextual effects can be calculated as the difference between L1- and L2-effects (see also Appendix A in the supplemental materials for a more detailed discussion of the use of model fit indices).

In the total sample, two models were fit: (a) a configural invariance model, in which a two-factor model for ASC and FUT was assumed to hold at both L1 and L2, but factor loadings were allowed to vary between the levels; and (b) a factorial invariance model in which factor loadings were restricted to be equal between L1 and L2. Indeed, fit indices were as good or better in the more restrictive model (no invariance: comparative fit index [*CFI*] = 0.979, Tucker-Lewis index [*TLI*] = 0.972, *RMSEA* = 0.010; invariance of factor loadings: *CFI* = 0.980, *TLI* = 0.976, *RMSEA* = 0.009).

Measurement invariance of ASC and FUT across the 56 countries was tested with a series of increasingly restrictive multigroup multilevel SEMs: (a) a configural invariance model in which a measurement model with two correlated factors for ASC and FUT was assumed to hold in all countries but no restriction on the factor loadings was imposed across levels and between countries; (b) the same model, with factor loadings restricted to be invariant across L1 and L2 within countries but freely varying across countries; (c) a model in which factor loadings were additionally restricted to be invariant across both L1 and L2 as well as across the 56 countries included in the analyses. The final model converged, but it resulted in estimated negative residual variances on L2 in some countries. Holding all of the residual variances at L2 invariant across countries solved the problem and resulted in a good model fit ($\chi^2 = 94{,}685.1$, $df = 5246$, *CFI* = 0.964, *TLI* = 0.965, *RMSEA* = 0.049). The resulting country-specific intraclass correlation coefficients for the latent variables ASC and FUT and for ACH are given in Table B1 in the supplemental materials.

Contextual effects models. Next, we sought to test the prediction of the BFLPE, negative contextual effects of ACH for both ASC and FUT using doubly latent multilevel SEMs (Marsh, Lüdtke, et al., 2009). The BFLPE model in the international sample included individual achievement (L1-ACH), school-average achievement (L2-ACH), and squared individual achievement (L1-ACH^2) as predictors. The last was included in line with previous BFLPE research (Marsh & Hau, 2003; Marsh & Rowe, 1996; Seaton et al., 2009) to control for nonlinearities in the relation of L1-ACH and ASC. All models included a random intercept for ASC and FUT but no further random effects.

All models were specified using implicit group-mean centering (Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995) of L1-ACH, which is the default for the latent aggregation procedure in Mplus. This implies that the regression coefficient of L2-ACH is not a direct estimate of the contextual effect but of the L2 effect

(see Enders & Tofighi, 2007; Kreft et al., 1995; Snijders & Bosker, 1999) that does not control for the effect of L1-ACH. An estimate of the contextual effect, the BFLPE, was obtained by subtracting the effect of L1-ACH from the effect of L2-ACH and obtaining its standard error with the multivariate delta method (Marsh, Lüdtke, et al., 2009; Raykov & Marcoulides, 2004; also see Appendix A in the supplemental materials). L1-ACH^2 was modeled by group-mean centering individual ACH and using the squared value as predictor in the model. Including L1-ACH^2 as a predictor in the model implied that the contextual effect changed with L1-ACH and that the parameter estimate for the contextual effect represents the BFLPE for a student whose achievement is average in his classroom (see the derivations in the supplemental materials, Appendix A).

To test the generalizability of the BFLPE using the international comparison sample, the BFLPE models were fit simultaneously in the 56 countries using a multigroup multilevel SEM in Mplus (Muthén & Muthén, 2010), fixing the factor loadings of the outcome variable (ASC and FUT, respectively) to be equal across L1 and L2 and across the countries. First, unconstrained models were estimated in which the parameters of the structural models were free to vary across the countries. These models were then compared to models in which the contextual effect of L2-ACH was held invariant across all 56 countries.

In all analyses, effect sizes were calculated according to the recommendations of Marsh, Lüdtke, et al. (2009) by the following formula:

$$\Delta = 2 \times \beta \times \sigma_{pred}/\sigma_y,$$

where β is the unstandardized regression coefficient, σ_{pred} is the standard deviation of the predictor variable (L1-ACH or L2-ACH), and σ_y is the standard deviation of the outcome variable (ASC or FUT, respectively). This effect size is comparable to Cohen's *d* (Cohen, 1988).

Multilevel mediation model. To test whether the effects of ACH on FUT were mediated by ASC, a multilevel mediation model (Preacher et al., 2010) was specified. For the total sample, this model used L1-ACH, L1-ACH^2, and L2-ACH as predictors for both ASC and FUT, thus replicating the structure of the basic contextual effects models. In addition, L1-ASC and L2-ASC were used as predictors for FUT. As a consequence, the paths from ACH to FUT no longer represented the total effects of L1-ACH and L2-ACH but rather the direct effects of L1-ACH and L2-ACH on FUT when ASC at both levels was controlled (e.g., MacKinnon, 2008). Again, due to the implicit group-mean centering of L1-ACH, an estimate of the direct contextual effect was obtained by subtracting the effect of L1-ACH from the effect of L2-ACH. The standard error for the direct contextual effect was obtained using the multivariate delta method (Raykov & Marcoulides, 2004).

The indirect effects of L1-ACH and L2-ACH on FUT that were mediated by ASC were obtained by multiplying the corresponding path coefficients from ACH to ASC and from ASC to FUT separately on L1 and L2. The indirect contextual effect on FUT was obtained by subtracting the indirect effect of L1-ACH from the indirect effect of L2-ACH. Standard errors for the indirect effect estimates were obtained with the multivariate delta method (Raykov & Marcoulides, 2004).

As a quadratic effect of L1-ACH on ASC was included, the indirect effect of L1-ACH on FUT and the indirect contextual effect varied as a function of L1-ACH (Hayes & Preacher, 2010). The unconditional indirect effects as obtained above reflect the indirect effects when L1-ACH is zero, corresponding to the grand mean in our analysis. In line with the rules outlined by Hayes and Preacher (2010), instantaneous indirect effects were obtained for both L1-ACH and the contextual effect of L2-ACH that showed changes in the indirect effects as functions of L1-ACH (for a detailed derivation, see supplemental materials, Appendix A).

Country-level moderators. To determine whether the total and direct effects of L1-ACH and the total and direct contextual effect of L2-ACH were moderated by country-level predictors, the parameter estimates of the unrestricted multigroup mediation model were related to the socioeconomic development and the five cultural dimensions from Hofstede et al. (2010) in a canonical correlation analysis. A canonical correlation analysis was the most appropriate method to explore these relations, as it deals with the multicollinearity of the country-level predictors and the effects in the mediation model and summarizes the complex multivariate relations between them. It would not have been possible to implement the alternative (a three-level SEM with random effects) because of prohibitive demands for computational resources and limitations of current software implementations. All variables and parameter estimates were standardized prior to the analyses. Significance of the canonical correlations was tested with Bartlett's chi-square test (Mardia, Kent, & Bibby, 1979) as implemented in the R-package yacca (Butts, 2009).

Results

Individual-Level and Contextual Effects for Academic Self-Concept in Science (ASC)

Is the contextual effect of attending a school with a high school-average ability negative for ASC in the total sample? In the total international sample of PISA 2006, L1-ACH was positively related to ASC ($\beta = 0.171$, $SE = 0.006$). In addition, there was evidence for a quadratic relation between L1-ACH and ASC ($\beta = 0.043$, $SE = 0.004$), indicating that higher achieving students tended to have higher ASC relative to what was implied by the linear effect. Most importantly for the present investigation, the contextual effect of L2-ACH, the BFLPE, was negative and large ($\beta = -0.205$, $SE = 0.014$). When individual differences in ACH were controlled, school-average ACH had a negative effect on ASC. The effect size of the BFLPE was −0.404, indicating a medium-sized effect. The presence of the positive quadratic effect at L1 implied that the BFLPE was larger (i.e., more negative) for students with relatively higher achievement compared to their schoolmates. All coefficients from this model are presented in Table 1. The resulting contextual effect is presented in the conceptual diagram in Figure 1 (upper panel).

Do these effects generalize across the 56 countries participating in PISA 2006? To test the generalizability of the total sample results to the 56 countries, the BFLPE model was fit to the multigroup sample. The resulting country-specific coefficients are given in Table B2 in the supplemental materials (Appendix B). The multigroup model fit the data well ($\chi^2 = 45897.5$, $df = 2733$, $CFI = 0.968$, $TLI = 0.969$, $RMSEA = 0.047$). Overall, the

Table 1
Coefficients of the BFLPE Models for the Complete PISA 2006 Dataset

Parameter	Self-concept		Career aspirations	
	Estimate	*SE*	Estimate	*SE*
Fixed effects				
Intercept	0	0	0	0
L1 science achievement	0.171	0.006	0.098	0.007
Quadratic L1 science achievement	0.043	0.004	0.052	0.006
Contextual effect science achievement	−0.205	0.014	−0.191	0.015
Variance components				
Intercept	0.026	0.002	0.049	0.003
L1 residual	0.373	0.008	0.570	0.008

Note. The intercept was fixed to zero for identification purposes. L1 = student level; L2 = school level; BFLPE = big-fish-little-pond effect.

findings from the total sample generalized well over the 56 countries participating in PISA 2006. In total, the BFLPE was negative and significantly different from zero in 50 countries and not significantly different from zero in the remaining six countries. A model in which the country-specific BFLPEs were restricted to be equal fit significantly worse ($\Delta\chi^2 = 514.6$, $df = 55$, $p < .001$), indicating that there was variability in the BFLPE across countries. However, differences in fit indices were essentially zero (*CFI* = 0.968, *TLI* = 0.969, *RMSEA* = 0.047, in the restricted model), indicating that the drop-off in fit, though statistically significant, was only marginal. The country-specific BFLPE effect sizes that express the size of the BFLPE relative to the country-specific standard deviations of ASC and ACH had a range from 0.070 to −0.673 with a mean of −0.282 and a standard deviation of 0.143. The country-specific BFLPE effects and their 95% confidence intervals are presented in Figure 2 (left panel).

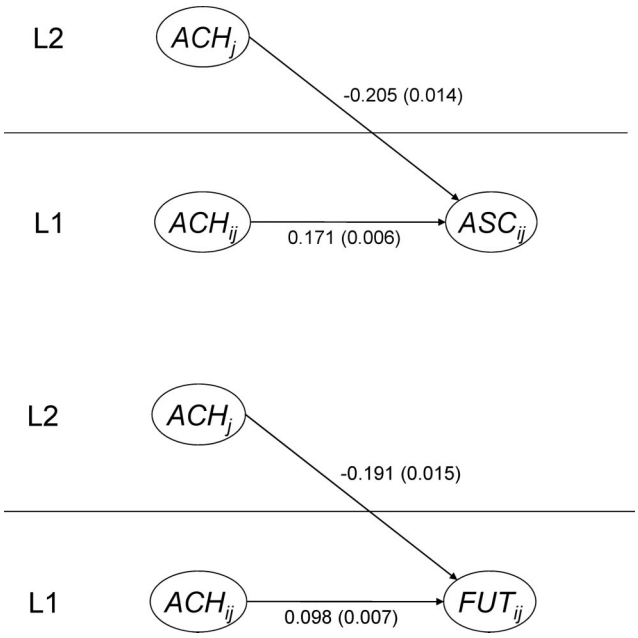

Figure 1. Conceptual path diagrams of the structural models of the big-fish-little-pond effect (BFLPE) of achievement (ACH) on academic self-concept in science (ASC, upper panel) and future career aspirations (FUT, lower panel). The path crossing the levels represents the contextual effect of school-level achievement on the individual-level outcome, the BFLPE. L1 = student level; L2 = school level. Note: The quadratic effects of ACH at L1 are not displayed in the path diagram.

In general, L1-ACH was positively related to ASC: Fifty-two countries showed a significant positive effect, while in three countries the effect of individual achievement was not significantly different from zero. Only one country (Indonesia) showed a significant negative effect. The effect sizes of the regression coefficients for L1-ACH ranged from −0.143 to 0.872, with a mean of 0.420 and a standard deviation of 0.234. The country-specific effects of L1-ACH on ASC and their 95% confidence intervals are presented in Figure 3 (left panel). Finally, the quadratic effect of L1-ACH was positive and significantly different from zero in 42 countries and not significantly different from zero in the remaining 14 countries. This implied that in the majority of the countries, higher achieving students were more strongly affected by the BFLPE.

Discussion. The findings from the first set of analyses replicated the BFLPE of ACH on ASC in the largest sample ever used to test it and further underline its status as a universal theory with strong cross-cultural generalizability (Seaton et al., 2009). Being schooled in a high-ability school reduced students' ASCs in science significantly in 50 of the 56 countries when individual ASC was controlled. We replicated the findings by Marsh and Hau (2003) and Seaton et al. (2009), who had provided evidence for the BFLPE with previous waves of PISA and for different subject areas (general and mathematics ASC). Taken together, these three studies show that the BFLPE generalizes not only across countries but also across three different subject areas. We now turn to the generalizability of the BFLPE to another important outcome variable: future plans and career aspirations.

Individual-Level and Contextual Effects for Career Aspirations in Science

Does the positive effect of L1-ACH and the negative contextual effects of L2-ACH generalize to FUT? To test the generalizability of the negative contextual effect of L2-ACH, we tested a parallel set of multilevel models for FUT rather than ASC. The pattern of results obtained from these models was very similar to the results for ASC. In the total sample, L1-ACH was positively

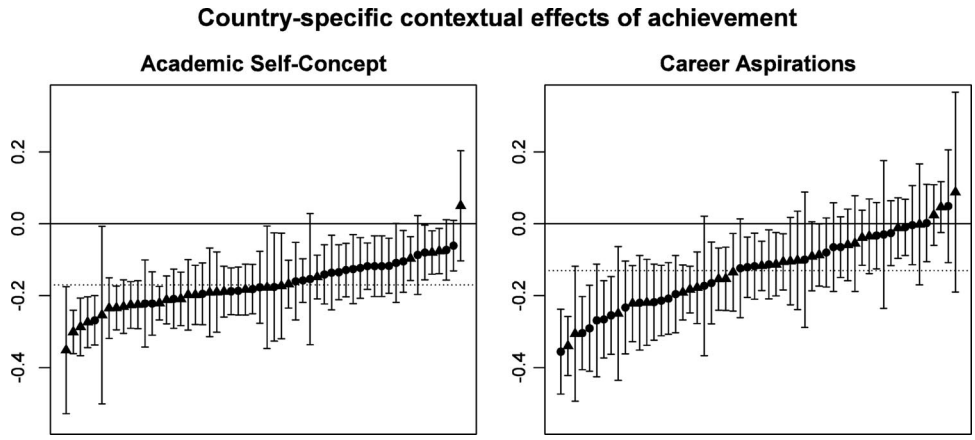

Figure 2. Caterpillar plots of country-specific contextual effects of science achievement on academic self-concept in science (left panel) and career aspirations in science (right panel). Organisation for Economic Co-operation and Development (OECD) countries are represented with triangles, non-OECD countries with circles. The dotted lines represent the average contextual effect over all countries. The error bars represent country-specific 95% confidence intervals.

related to FUT ($\beta = 0.098$, $SE = 0.007$). The quadratic effect of L1-ACH was also clearly significant ($\beta = 0.052$, $SE = 0.006$), showing that the relation between L1-ACH and FUT became steeper for high-achieving students. Again, a negative contextual effect emerged ($\beta = -0.191$, $SE = 0.015$). Students with the same ACH level expressed lower FUT when attending schools with higher average ACH levels. In the appropriate effect size metric, the contextual effect of school-average ACH was −0.311, indicating a medium-sized effect that was nearly as large as the negative effect of school-average ACH on ASC (−0.404) already considered. Again, the presence of a positive quadratic effect of L1-ACH implied that the BFLPE was slightly larger for students with relatively higher ACH compared to their schoolmates. All coefficients from this model are given in Table 1. The resulting relations between the variables are presented in Figure 1 (bottom panel).

Do these effects generalize across the 56 countries that participated in PISA 2006? To test for the generalizability of the negative contextual effect of ACH on FUT across the 56 countries of the international sample, the model that included the main effects of L1-ACH and L2-ACH as well as quadratic L1-ACH was fit in the multigroup data set. The model fit the data well ($\chi^2 = 48635.8$, $df = 1281$, $CFI = 0.956$, $TLI = 0.954$, $RMSEA = 0.072$).

The contextual effect of L2-ACH on FUT was negative and significantly different from zero in 31 of the 56 countries. A model that restricted the BFLPE across the countries to be equal fit the data significantly worse ($\Delta\chi^2 = 725.6$, $df = 55$, $p < .001$).

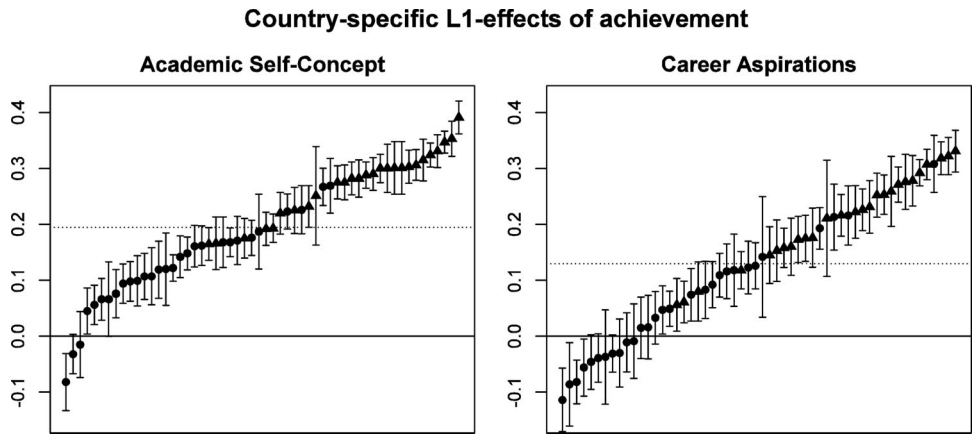

Figure 3. Caterpillar plots of country-specific L1-effects of science achievement on academic self-concept in science (left panel) and career aspirations in science (right panel). Organisation for Economic Co-operation and Development (OECD) countries are represented with triangles, non-OECD countries with circles. The dotted lines represent the average L1-effect over all countries. The error bars represent country-specific 95% confidence intervals.

However, the fit indices in the restricted model were as good as or better than the fit indices of the original models (*CFI* = 0.955, *TLI* = 0.955, *RMSEA* = 0.071), indicating that the drop-off in fit—though statistically significant—was very small, and the BFLPEs were reasonably consistent across the countries. In the effect size metric that expresses the size of the contextual effects relative to the country-specific standard deviations of FUT and ACH, the contextual effects ranged from 0.105 to −0.502, with a mean of −0.167 and a standard deviation of 0.145. The country-specific contextual effects and their 95% confidence intervals are plotted in Figure 2 (right panel) and can also be found in Table B3 in the supplemental materials (Appendix B).

The effect of L1-ACH on FUT was positive and significantly different from zero in 42 of the 56 countries and significantly negative in four countries (Colombia, Kyrgyztan, Montenegro, and Romania). The country-specific effect sizes for the effect of L1-ACH ranged from −0.164 to 0.670, with a mean of 0.229 and a standard deviation of 0.221. Figure 2 (right panel) shows the country-specific effects of L1-ACH and their 95% confidence intervals. The quadratic effects of L1-ACH were positive and significant in 42 of the 56 countries, showing substantial nonlinearities in the relation of ACH and FUT. High-achieving students (relative to the average of the school they were attending) consistently reported higher career intentions than would be expected from a simple linear trend. This also affected the BFLPE for high-achieving students in these countries: In 17 of the 25 countries where the BFLPE had not been statistically significant for the average student in a school, there was a positive quadratic effect, indicating that the BFLPE was affecting the career aspirations of higher performing students more strongly.

Discussion. In line with the findings for ASC, there was a strong and significant negative contextual effect of L2-ACH on FUT in the total sample. Students who attended higher ability schools reported fewer intentions of pursuing a science-related career than their equally able peers who attended a lower ability school. This effect generalized to a majority of countries (31) in the sample. However, the average BFLPE and its generalization were not as strong as with the BFLPE for ASC. The importance of school-average ACH for ASC seems to be larger, possibly because of the higher relevance of the frame of reference of other students for the assessment of one's own ability compared to the development of career aspirations. However, the quadratic effect of ACH implied that although the BFLPE for an average student in a school might be somewhat smaller for FUT, the BFLPE still adversely affected relatively higher performing students.

Similarly, the results of the L1 relations between ACH and FUT revealed that high-achieving students also reported more science-related career intentions than their lower performing peers in the majority of countries (42). This relation was quadratic in many countries (42), implying that career intentions did not rise equally with increases in achievement across the whole achievement spectrum. Rather, the positive quadratic effects indicated that the relation between ACH and FUT was flatter for students with relatively lower achievement than for students with higher achievement. Students only considered taking up a science-related career if they were reasonably high-achieving already. Once their ACH was above a certain threshold, however, even small increases in ACH made it more likely that they were considering a career in a science-related field but also that they were more negatively affected by being in a selective schooling environment.

Mediation Model for FUT

Are there indirect effects of ACH on FUT mediated by ASC? In the next step, we explored how the effects of L1-ACH and L2-ACH on FUT were mediated by ASC. For this purpose, we fit a multilevel mediation model (Pituch & Stapleton, 2011; Preacher et al., 2010), first to the total sample and then within each country separately, using a multigroup multilevel SEM. The parameter estimates and standard errors for the total international sample are given in Table 2 and displayed in Figure 4. We present the resulting mediational chains in Figure 5.

We first consider the relations between ASC and FUT (controlling for ACH) that appeared for the first time in the mediation model. The effect of L1-ASC on L1-FUT was positive ($\beta = 0.600$, $SE = 0.007$), indicating that students with higher ASC expressed higher intentions of taking up a science-related career. The contextual effect of ASC on FUT was positive ($\beta = 0.250$, $SE = 0.078$). Being in a school environment with other students who perceive their abilities in science to be high added somewhat to one's own career ambitions in science, over and above the effect of individual self-concept. However, in the effect size metric, this coefficient was only 0.101, indicating that the additional contextual effect of L2-ASC on FUT was small.

Next, we turn to the effects of ACH on FUT at L1 and L2 after controlling for the effects of ASC. Taking the effects of ASC into account substantially reduced the effect of L1-ACH on FUT; the linear effect of L1-ACH on FUT ($\beta = -0.005$, $SE = 0.006$) was no longer statistically significant. The indirect effect of L1-ACH mediated by L1-ASC was large in absolute size ($\beta = 0.103$, $SE = 0.004$). The κ^2 statistic (Preacher & Kelley, 2011) that expresses the size of the indirect effect relative to the maximum possible indirect effect, given the constraints of the variance-covariance matrix of the involved variables, was $\kappa^2 = 0.113$ for the L1 mediation effect. In addition, there was still a quadratic direct effect of L1-ACH on FUT ($\beta = 0.026$, $SE = 0.005$) indicating that higher achieving students were more likely to express higher science-related career intentions, even after differences in their ASC were taken into account.

The direct contextual effect of L2-ACH was also reduced compared to the model without ASC (see Table 2). However, in contrast to the effect of L1-ACH, it remained statistically significant ($\beta = -0.056$, $SE = 0.014$). This resulted in an effect size of −0.091 for the direct contextual effect, compared to an effect size of −0.304 of the total contextual effect in the unmediated model. The indirect contextual effect ($\beta = -0.135$, $SE = 0.010$) accounted for over 70% of the total contextual effect of L2-ACH on FUT. The κ^2 statistic of the indirect contextual effect was equal to 0.110 (see the supplemental materials for details of the calculation of κ^2 for the indirect contextual effect).

The model relating ASC to L1- and L2-ACH—and the resulting effects—were essentially the same in the mediation model. They included significant linear and quadratic effects for L1-ACH, as well as a significant L2-effect of ACH, and a negative contextual effect of L2-ACH, the BFLPE on ASC. The presence of the quadratic effect of L1-ACH on ASC implied, in line with the derivations by Hayes and Preacher (2010), that the indirect effects

Table 2
Coefficients of the BFLPE Multilevel Mediation Model for the Complete PISA 2006 Dataset

Parameter	Estimate	*SE*
Fixed effects		
Directly estimated[a]		
Career aspiration		
L1 Science achievement (direct effect)	−0.005	0.006
Quadratic L1 Science achievement	0.026	0.005
L1 Academic self-concept in science	0.600	0.007
L2 Science achievement (direct effect)	−0.061	0.012
L2 Academic self-concept in science	0.850	0.077
Academic self-concept in science		
L1 Science achievement	0.172	0.006
Quadratic L1 Science achievement	0.043	0.003
L2 Science achievement	−0.037	0.012
Derived from estimated effects[b]		
Career aspiration		
L1 Science achievement (total effect)	0.098	0.007
L1 Science achievement (indirect effect)	0.103	0.004
Instantaneous indirect effect coefficient for L1 Science achievement	0.052	
L2 Science achievement (total effect)	−0.093	0.013
L2 Science achievement (indirect effect)	−0.031	0.010
Contextual effect of L2 Science achievement (total effect)	−0.191	0.015
Contextual effect of L2 Science achievement (direct effect)	−0.056	0.014
Contextual effect of L2 Science achievement (indirect effect)	−0.135	0.011
Contextual effect of academic self-concept in science	0.250	0.078
Instantaneous indirect effect coefficient for contextual effect of L2 Science achievement	−0.052	
Academic self-concept in science		
Contextual effect of L2 Science achievement	−0.209	0.014
Variance components		
Career aspirations		
Intercept	0.029	0.002
L1 Residual	0.434	0.006
Academic self-concept in science		
Intercept	0.027	0.002
L1 Residual	0.381	0.008

Note. L1 = student level; L2 = school level. BFLPE = big-fish-little-pond effect.
[a] See also Figure 4. [b] See also Figure 5.

of L1-ACH on FUT and the contextual effect of L2-ACH on FUT depended on L1-ACH. The coefficient of the instantaneous indirect effects at L1 was 0.052. This implied that the indirect effect for relatively high-achieving students was stronger than for relatively low-achieving students. Therefore, having a higher self-concept in science was more important for bringing about career intentions in high-achieving than in low-achieving students. However, ASC was also more important in transmitting the negative contextual effect of school-average ACH on high-achieving students' career intentions in science. The higher the students' individual ACH was relative to their school, the more negative was the contextual indirect effect Their career intentions suffered relatively more from the BFLPE as transmitted by ASC.

A breakdown of the indirect contextual effect into the L2- and the L1-mediating chain (see Figure 5 for a visualization and supplemental Appendix A for a derivation) showed that L1-ASC was relatively more important for bringing about the indirect contextual effect: The indirect effect component based on the L1-mediating chain was equal to −0.126 (*SE* = 0.009). The indirect effect component based on the L2-mediating chain, in contrast, was equal to −0.009 (*SE* = 0.004). This suggested that the mediation of the contextual effect of ACH on FUT mostly took place at L1; that is, that individual ASC was by far more important than L2-ASC for transmitting the negative contextual effect of L2-ACH.

How do the effects in the mediation model generalize across the countries participating in PISA 2006? We tested whether the mediation of the negative contextual effect of L2-ACH and the positive effect L1-ACH on FUT by ASC generalized over the 56 countries. The indirect contextual effect of L2-ACH on FUT was significant in 34 of the 56 countries. The effect sizes of the indirect contextual effects ranged from −0.450 to 0.318, with a mean of −0.132 and a standard deviation of 0.108. As there were significant positive quadratic effects of L1-ACH on ASC in 41 countries, the indirect contextual effect was stronger (i.e., more negative) for relatively high-achieving students in these countries. This was the case for 15 out of the 22 countries where the indirect contextual effect for an average student in a school was not statistically significant.

Without controlling for the effects of ASC, the contextual effect of L2-ACH on FUT had been negative and significantly different from zero in 31 countries. Including ASC in the multilevel medi-

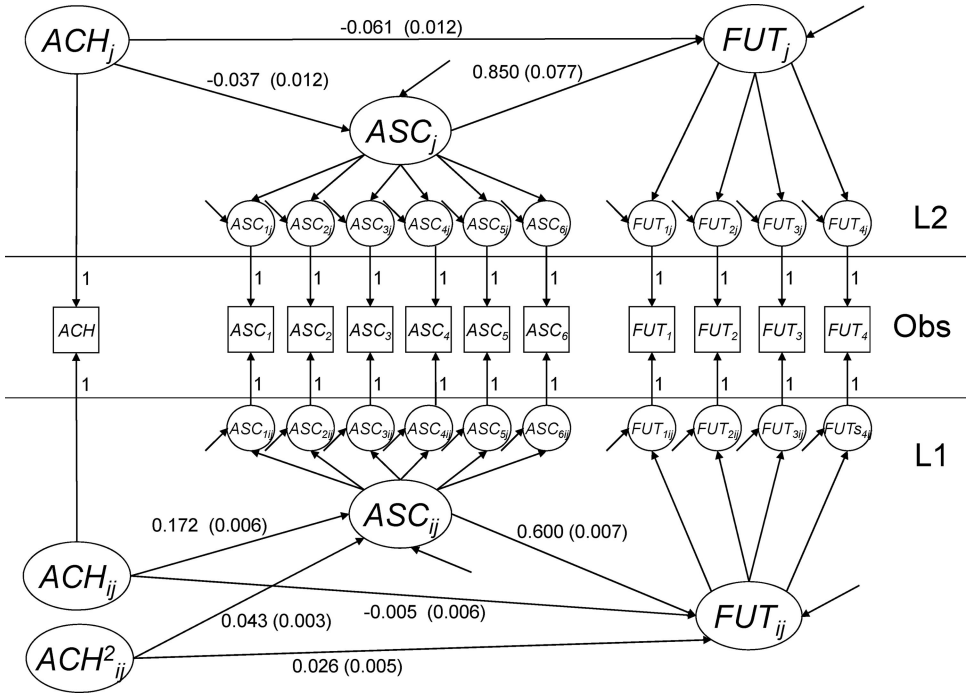

Figure 4. Path diagram of the multilevel mediation model for the total sample. The model is based on implicit group-mean centering of the L1 variables. Paths at L1 represent L1-effects; paths at L2 represent L2-effects. Contextual effects have to be derived from these parameters. ACH = science achievement; ASC = academic self-concept in science; FUT = career aspirations in science; L1 = student level; L2 = school level; Obs = observed variables.

ation model, this number was reduced to 15. Three countries showed a significant positive direct contextual effect. Overall the effect sizes of the direct contextual effects of L2-ACH after controlling for ASC ranged from 0.171 to −0.451, with a mean of −0.036 and a standard deviation of 0.117. The country-specific direct and indirect contextual effects are also given in Table B4 in the supplemental materials (Appendix B).

How did the mediation effect of the L1 relation between ACH and FUT by ASC generalize across the 56 countries? The effect sizes of the indirect effects of L1-ACH on L1-FUT (mediated by L1-ASC) ranged from −0.064 to 0.373, with a mean of 0.146 and a standard deviation of 0.086. They were positive and significant in 54 countries and nonsignificantly different from zero in two countries (Kyrgyzstan and Thailand). This finding implied that in almost all of the countries, L1-ASC transmitted at least some of the positive effect of L1-ACH on FUT. The effect sizes of the direct effect of L1-ACH on FUT after controlling for ASC ranged from −0.164 to 0.186, with a mean of 0.014 and a standard deviation of 0.082. Without controlling for the effects of ASC, the effect of individual ACH on FUT was positive and significant in 42 countries. In the multilevel mediation model, only 20 countries had a positive and signifi-

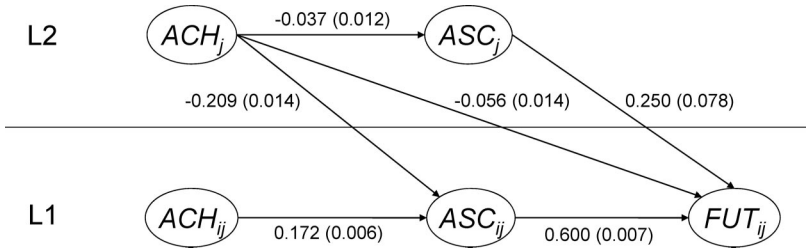

Figure 5. Conceptual path diagram of the multilevel mediation model representing the relation of achievement (ACH), academic self-concept in science (ASC) and future career aspirations (FUT). Note: The quadratic effects of ACH at L1 are not displayed in the path diagram. The direct effect of ACH on FUT on L1 was not statistically significant and is not shown in the path diagram. The paths at L1 represent L1-effects, the path at L2 represents L2-effects, and the paths crossing the levels represent contextual effects. ACH = science achievement; ASC = academic self-concept in science; FUT = career aspirations in science; L1 = student level; L2 = school level.

cant direct effect of individual ACH on FUT, while 10 countries had a negative and significant direct effect. In the remaining 26 countries, the relation between L1-ACH and FUT was no longer statistically significant.

Discussion. The multilevel mediation model shed further light onto the processes that brought the effects of ACH on FUT about. After controlling for ASC, the direct effect of L1-ACH on FUT was essentially zero in the total sample. Similarly, the negative contextual effect of L2-ACH on FUT was substantially reduced when ASC was controlled. Again, these effects generalized reasonably over the 56 separate countries. There were significant indirect effects at L1 and significant indirect contextual effects in the majority of the countries. The direct effects of L1-ACH and the contextual direct effect of L2-ACH were substantially reduced in the majority of nations participating in PISA 2006, attesting to the generalizability of the mediating role of ASC in bringing about motivation and career aspirations.

Relation With Country-Level Moderators

Can the differences in country-specific effects of ACH be related to country-level moderators such as country-level cultural orientations and socioeconomic development? To explore the relation of country-level variables to the country-specific effects in the multilevel mediation model, we conducted a canonical correlation analysis. The country-level variables (socioeconomic development as measured by the U.N. Human Development Index and the values on the five cultural orientations: individualism, power distance, masculinity, long-term orientation, and uncertainty avoidance) formed one set of variables in the canonical correlation analysis; the country-specific effects of L1-ACH (direct effect on ASC, direct and total effect on FUT) and the contextual effects of L2-ACH (direct effect on ASC, direct and total effect on FUT) formed the other set of variables. Nine countries with missing data on the cultural orientation measures were excluded from these

Table 3
Results of the Canonical Correlation Analysis

	Component 1			Component 2			Component 3			
Variable/Statistic	B[a]	r[b]	r^2[c]	B[a]	r[b]	r^2[c]	B[a]	r[b]	r^2[c]	h^2[d]
Cultural factors										
Human Development Index	−0.91	−0.99	0.98	0.41	0.00	0.00	0.67	0.04	0.00	0.98
Power distance	−0.05	0.68	0.47	−0.29	0.17	0.03	−0.07	−0.04	0.00	0.50
Individualism	−0.02	−0.71	0.50	−0.99	−0.49	0.24	−1.09	−0.49	0.24	0.98
Masculinity	0.03	0.00	0.00	−0.07	−0.15	0.02	−0.01	−0.15	0.02	0.04
Uncertainty avoidance	0.12	0.44	0.19	0.06	0.20	0.04	−0.26	0.02	0.00	0.23
Long-term orientation	−0.05	−0.08	0.01	0.78	0.73	0.53	−0.65	−0.67	0.44	0.98
Adequacy			0.36			0.14			0.12	
Redundancy (R_{X^*xs})			0.31			0.05			0.03	
Squared canonical correlation R_c		0.92	0.85		0.56	0.32		0.49	0.24	
Country-specific effects										
Adequacy			0.42			0.18			0.18	
Redundancy (R_{Y^*xs})			0.36			0.06			0.04	
ASC on L1-ACH	−1.07	−0.99	0.98	0.70	−0.08	0.01	−0.88	−0.05	0.00	0.99
FUT on L1-ACH (direct)	−0.39	−0.72	0.53	1.47	0.01	0.00	0.34	0.54	0.29	0.82
FUT on L1-ACH (total)	0.38	−0.91	0.84	−1.97	−0.10	0.01	0.99	0.33	0.11	0.96
ASC on L2-ACH	−0.02	0.37	0.14	0.05	0.65	0.43	0.58	0.25	0.06	0.63
FUT on L2-ACH (direct)	−0.04	−0.11	0.01	−0.71	0.22	0.05	−0.18	−0.64	0.41	0.47
FUT on L2-ACH (total)	−0.03	−0.12	0.01	1.29	0.76	0.57	−0.41	−0.44	0.20	0.78

	Significance tests of the canonical correlations using Bartlett's χ^2 test			
Canonical variate	ρ^2	χ^2	df	p
1	0.855	108.030	36	<0.001
2	0.318	31.830	25	0.163
3	0.243	16.685	16	0.406
4	0.120	5.673	9	0.772
5	0.013	0.625	4	0.960
6	0.002	0.092	1	0.761

Note. ASC = academic self-concept in science; L1-ACH = student-level achievement in science; FUT = career aspirations in science; L2-ACH = school-level achievement in science; direct = direct effect; total = total effect; ρ^2 = squared canonical correlation; χ^2 = χ^2 test statistic; df = degrees of freedom; p = p value of the test that the current and the remaining canonical variates are significant. All effects at L2-variables are contextual effects.
[a] Standardized function coefficients. [b] Structure coefficients (when associated with the variables), or canonical correlation coefficients (when not associated with the variables). [c] Squared structure coefficients (when associated with the variables); adequacy and redundancy coefficients, and squared canonical correlation coefficients (when not associated with the variables). [d] Communality: % variance in a variable jointly extracted by the two canonical functions.

supplemental analyses alone, resulting in a total sample of 47 countries. Both the country-level predictors and the country-specific effects were highly correlated (see the correlation matrix of the resulting sample in the supplemental materials, Appendix B, Table B5), and a canonical correlation analysis was deemed appropriate to summarize the dimensionality and interrelation of the two sets of variables.

Tests of dimensionality for the canonical correlation analysis using Bartlett's χ^2 test (as shown in Table 3) indicated that there was one canonical dimension that was statistically significant. Dimension 1 accounted for a canonical correlation of 0.925 between the two sets of variables. While Dimensions 2 and 3 also had relatively large canonical correlations (0.564 and 0.493), they were not statistically significant. The aggregate redundancies that measure the percentage of variance that is explained by all six canonical dimensions were 39.8% for the country-level predictors and 46.9% for the country-level effects. Table 3 presents the standardized canonical and structural coefficients for the first three dimensions across the two sets of variables. The canonical coefficients can be interpreted as partial regression coefficients of the components on the observed variables, whereas the structural coefficients (or loadings) represent first-order correlations between the canonical dimensions and the observed variables that are not controlled for the influences of other observed variables.

The first dimension was characterized by a large canonical coefficient for the U.N. Human Development Index and high structural coefficients for the Human Development Index, individualism (both negative), and power distance (positive) in the set of country-level predictors. On the set of country-specific effects, the three L1-effects had large canonical and structural coefficients for this dimension. Overall, the first canonical variate explained 35.7% of the variance in the country-level predictors and 41.8% of the variance in the country-specific effects. The redundancy coefficients that measure the explained variance across sets were 30.5% for the set of country-level predictors and 35.7% for the set of country-level effects, indicating that 76.7% and 76.2% of the total variance explained in the two variable sets was explained by the first canonical variate.

Discussion. The results from the canonical correlation analysis suggested that the L1-effects of ACH on both ASC and FUT were higher in more developed and individualistic countries. Both dimensions were highly correlated in the country sample ($r = .662$). This pattern was also evident in the loading pattern on the first canonical variate. There was a large canonical coefficient only for the U.N. Human Development Index. This implied that after controlling for correlations among the predictor variables, only the U.N. Human Development Index proved to be an important predictor. However, the large structural coefficients for both the U.N. Human Development Index and individualism, representing first-order relations between the respective variables and the canonical variates that are uncorrected for relations between the predictors, suggested that both variables were important. Differences in individual ACH seemed to translate into differences in ASC and FUT more strongly in highly developed and individualistic countries. These findings are in line with the results by Chiu and Klassen (2010). Using data from PISA 2003, they reported stronger effects of individual ACH on ASC in math in more developed countries. However, similar to the findings in the canonical correlation analysis, the effects of individualism on the L1-effects of ACH were not significant when developmental status was controlled.

Other relations between the two sets of variables in the canonical correlation analysis, in particular relations of the contextual effects of ACH with country-level predictors, were not statistically reliable. These findings are in line with the tests of the invariance of the contextual effects of ACH on ASC and FUT, which showed that these effects were reasonably similar across the 56 countries. They are also in line with the findings by Seaton et al. (2009), who found that the interaction between country-level individualism and the BFLPE on ASC for math was significant but very small and that there was no relation between developmental status and the BFLPE. Hence, these findings support the notion of universal generalizability of the BFLPE, the negative contextual effect of L2-ACH on both ASC and FUT.

General Discussion

This study is a substantive–methodological synergy (Marsh & Hau, 2007), extending previous cross-cultural research on the BFLPE in the following ways: (a) Our sample, taken from PISA 2006, included more countries and a more diverse sample of countries than any other previous BFLPE study; (b) ours is the first cross-cultural study using PISA data (and one of the few BFLPE studies) to focus on ACH and ASC in science (also see Nagengast & Marsh, 2011), a subject area with high policy relevance for developed and developing countries; (c) we showed that the BFLPE generalized to career aspirations in science and that these effects were mediated by ASC; and (d) we used multilevel SEMs, cutting-edge quantitative methods, to implement the contextual effects and the multilevel mediation models.

Including this study, there now have been three large-scale investigations (Marsh & Hau, 2003; Seaton et al., 2009) into the generalizability of the BFLPE across countries using PISA data. Although the subject focus changed between the studies (general ASC, Marsh & Hau, 2003; math ASC, Seaton et al., 2009; science ASC, this study) and the employed statistical models have become more refined over time, the results point to the broad generalizability of the BFLPE. Table 4 shows the comparison of country-specific BFLPE estimates from the three studies. It is striking that in 123 different samples, the BFLPE was estimated as negative in all but one analysis and proved to be statistically significant in 114 out of 123 cases. This underscores the broad generalizability of the BFLPE and the universality of negative effects on ASC and related outcomes due to ability segregation in the educational system.

Academic self-concept in science in particular and career aspirations more generally have not been the focus of BFLPE researchers. Previous research has focused on general academic self-concept or reading and math self-concept. While research into the development of career choices (e.g., Betz & Hackett, 1981; Hackett & Betz, 1981; Lent et al., 2001) has found support for the importance of self-related perceptions for the development of career choices in science and engineering subjects, our study is apparently the first that was not restricted to Western and developed countries to demonstrate the negative contextual influence of being schooled with other high-

Table 4
Comparison of Country-Specific BFLPE Estimates for Academic Self-Concept From Comparative Studies Using Data From PISA

Country	PISA 2006 Science	PISA 2003 Math	PISA 2000 General
Azerbaijan	−0.154		
Argentina	−0.177		
Australia	−0.168	−0.281	−0.23
Austria	−0.231	−0.483	−0.23
Belgium	−0.183	−0.447	−0.12
Brazil	−0.118	−0.372	−0.26
Bulgaria	−0.073		
Canada	−0.234	−0.427	
Chile	−0.118		
Chinese Taipei	−0.080		
Colombia	−0.129		
Croatia	−0.123		
Czech Republic	−0.221	−0.446	−0.24
Denmark	−0.190	−0.296	−0.17
Estonia	−0.182		
Finland	−0.254	−0.301	−0.14
France	−0.226	−0.383	
Germany	−0.301	−0.713	−0.30
Greece	−0.148	−0.174	
Hong Kong	−0.209	−0.200	
Hungary	−0.209	−0.323	−0.05
Iceland	−0.173	−0.209	−0.18
Indonesia	−0.195	−0.235	
Ireland	−0.191	−0.103	−0.24
Israel	−0.222		
Italy	−0.212	−0.409	−0.36
Japan	−0.097	−0.307	
Jordan	−0.105		
Korea	0.050	−0.014	−0.02
Kyrgyzstan	−0.187		
Latvia	−0.118	−0.221	−0.06
Liechtenstein		−0.554	−0.20
Lithuania	−0.135		
Luxembourg	−0.076	−0.428	−0.17
Macao-China	−0.160	−0.330	
Mexico	−0.061	−0.357	−0.08
Montenegro	−0.136		
Netherlands	−0.287	−0.696	−0.26
New Zealand	−0.235	−0.314	−0.26
Norway	−0.198	−0.168	−0.18
Poland	−0.126	−0.279	
Portugal	−0.274	−0.205	−0.18
Qatar	−0.269		
Romania	−0.087		
Russian Federation	−0.222	−0.187	−0.21
Serbia	−0.141	−0.181	
Slovak Republic	−0.189	−0.411	
Slovenia	−0.188		
Spain	−0.080	−0.244	
Sweden	−0.177	−0.202	−0.33
Switzerland	−0.198	−0.446	−0.17
Thailand	−0.176	−0.194	
Tunisia	−0.117	−0.161	
Turkey	−0.109	−0.252	
United Kingdom	−0.225	−0.344	−0.23
United States	−0.352	−0.230	−0.26
Uruguay	−0.158	−0.240	

Note. Results for PISA 2006 are taken from the present study; results for PISA 2003 are taken from Seaton, Marsh and Craven (2009); and results for PISA 2000 are taken from Marsh and Hau (2003). BFLPE = big-fish-little-pond effect.

achieving peers on students' intentions to pursue a career in science. Taken together with the results for ASC and the mediation models, we not only found support for the generalizability of the BFLPE to a different content area but also demonstrated its far-reaching negative consequences on career aspirations and showed that these effects were mediated by ASC. While the negative contextual effect on career intentions was consistent across the countries in our study (as shown by the negligible change in model fit of the restricted model), graphical inspections suggested that it was somewhat stronger in more economically and socially developed countries. However, this suggestion was not supported by the canonical correlation analysis. Hence, further research into the mechanisms of contextual effects of achievement on career aspirations and its cultural boundary conditions are needed.

Overall, the relations between ACH and FUT were somewhat smaller in magnitude than the relations between ACH and ASC. As a direct consequence, the corresponding effects were statistically significant in a smaller number of countries in the multigroup analysis. One possible explanation for this difference might be that the coupling between ACH and FUT is comparatively looser and more complex. Although the linear relation between ACH and FUT was not as strong as the linear relation between ACH and ASC, there was evidence for (positive) quadratic effects that suggested that ACH differences became more relevant for developing career aspirations for high achieving students. Students will consider a career involving science only if they are relatively high in science achievement. These quadratic trends also influenced the BFLPE and the mediation effects: The BFLPE was stronger for high-achieving students, as were the indirect effects mediated by ASC. Hence, although the relations between ACH and FUT appeared to be smaller for average students in a school, the presence of quadratic effects suggests a richer picture.

While we were able to take advantage of the large and representative sample of students available in PISA 2006, applied researchers will typically conduct research on frame-of-reference effects within one country at a time. How do our findings (in particular, the fact that there were countries in which the BFLPE was not statistically significant) inform this kind of research? Obviously, our results attest to the generalizability of the BFLPE on ASC and FUT. Although they were not statistically significant in all countries, there was little variability in the country-specific BFLPEs, as shown by the minor changes in model fit when they were kept constant across the countries. The canonical correlation analysis showed that this remaining variability could not be reliably related to country-level moderators and that further research is needed to understand these country-level differences. The effects of ACH on ASC and FUT at the individual level, in contrast, were stronger in more developed and more individualistic countries. The presence of quadratic effects that complicated the interpretation of the BFLPE also suggests that researchers analyzing data from a single country have to take great care when specifying statistical models and interpreting their analyses. Even though the BFLPE for an average student might appear to be zero, the presence of quadratic effects indicates that some students' ASC or FUT might still suffer from it.

Limitations

Causality. The most crucial limitation of the current study lies in its cross-sectional nature, which makes causal interpretations of the findings difficult to justify, as the temporal ordering of the variables is unclear, reciprocal relations between the constructs are likely, and the results of cross-sectional mediation models do not necessarily capture longitudinal processes (Maxwell, Cole, & Mitchell, 2011). However, in the present situation, the case could be made for ACH as the most proximal variable in the causal chain that influences both ASC and FUT as outcome variables. Also, the ordering of ASC before FUT can be justified by pointing out that ASC as a construct is more strongly focused on perceptions of current academic accomplishments, while FUT has a stronger prospective, future-oriented component to it. Indeed, previous longitudinal studies on the BFLPE for FUT (Marsh, 1991: Marsh & O'Mara, 2010) with a clear temporal ordering of the three constructs have also found that ASC mediated the effects of ACH on FUT. However, the causal direction of the relation between these variables cannot be inferred from cross-sectional data alone.

In relation to ASC, there already is a large body of literature that addresses these concerns (see Marsh, 2006; Marsh, Seaton, et al., 2008). Quasi-experimental, longitudinal studies (e.g., Marsh et al., 2000) show that students' ASC declines when students shift from mixed-ability schools to academically selective schools over time (based on pre–post comparisons) and compared to students matched on academic ability who continue to attend mixed-ability schools. There is support for the BFLPE in studies where achievement is based on tests administered before students began high school (e.g., Marsh et al., 2000). Extended longitudinal studies (Marsh et al., 2000; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007) show that the BFLPE grows more negative the longer students attend a selective school and is maintained even 2 and 4 years after graduation from high school. Also, there is good support for the theoretical underpinnings of the BFLPE, as it is largely limited to academic components of self-concept and nearly unrelated to nonacademic components of self-concept and self-esteem (Marsh, 1987; Marsh & Parker, 1984). However, further longitudinal and intervention studies would be useful to bolster the case for mediation of the effects of L1- and L2-ACH on FUT by ASC.

In addition, the causal interpretation of simple linear mediation models has recently come under increased scrutiny even in the most favorable situation of a randomized experiment (Bullock, Green, & Ha, 2010; Foster, 2010; Imai, Keele, & Tingley, 2010; Sobel, 2008; VanderWeele, 2010). Extensions of alternative methods (Imai et al., 2010; Sobel, 2008) to multilevel mediation and truly observational studies are needed. However, the most promising avenue for establishing causality lies in an encompassing research program that disentangles the relations between hypothetical constructs using multiple methods and multiple longitudinal, cross-sectional and intervention studies.

Tests of indirect effects. We based our tests of the indirect effects on analytical standard errors obtained with the multivariate delta methods (Rao, 1973; Raykov & Marcoulides, 2004), as implemented in Mplus. In single-level mediation models, this standard error is also known as Sobel's (1982) test and is widely used. However, recent statistical derivations and simulation studies for single-level (MacKinnon, Lockwood, & Williams, 2004; Selig & Preacher, 2008) and multilevel mediation models (Preacher et al., 2010, 2011) suggest that the symmetric confidence interval of the indirect effect that is implied by this method is not appropriate. Alternative methods for assessing the confidence limits are resampling methods like the bias corrected (Preacher & Hayes, 2004) and the parametric bootstrap (Bauer, Preacher, & Gil, 2006; MacKinnon et al., 2004). Given the large sample sizes and computational demands of our analyses, it was infeasible to implement a full bias-corrected bootstrap procedure to test the indirect effects in the multilevel mediation models. However, we implemented the parametric bootstrap suggested by Bauer et al. (2006) and MacKinnon et al. (2004) using Selig and Preacher's (2008) implementation to test the robustness of our conclusions for the total sample. The results were essentially the same.

Three-level models. A further limitation of the statistical analyses was that we could not explicitly model the three-level structure of the data in a three-level SEM. Although three-level SEMs can be estimated in GLLAMM (Skrondal & Rabe-Hesketh, 2004) and WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), they are currently not implemented in the version of Mplus (Muthén & Muthén, 2010) that we used in our analysis. As the computational demands for implementing such a model with the alternative programs would have been prohibitive, given the large samples we were working with, we implemented the following measures to deal with the third level of nesting. We centered all variables around their country mean, thus removing country-level variance, which would have been falsely attributed to the school-level (Moerbeek, 2004), from the total group analysis. We also included country as a stratification variable in the total group analysis, thus controlling standard errors and model fit statistics for the nesting of schools and students within countries. The multigroup models that were used to test the generalizability of effects across countries explicitly took the third-level into account and modeled interactions of the core parameters with the countries as fixed effects (compared to random slopes with an additional distributional assumption in a three-level SEM). Finally, the parameters obtained from these models were used in the canonical correlation analysis to study interrelations with country-level moderators. Nevertheless, the development of efficient implementations of multilevel SEMs with more than two levels and random effects remains an important area for further methodological research and software development (see also Preacher, 2011).

Implications and Conclusion

Our results showed that being schooled with other high-achieving students has a generalizable and consistent negative impact on ASC and on career aspirations in science—the BFLPE. One of the reasons for achievement differences between schools, apart from regional differences, is implicit or explicit tracking of students. Meta-analytical research on the effects of tracking (Hattie, 2002) reveals the BFLPE in relation to ASC but shows that tracking has only very small effects on achievement, particularly in comparison to typical effects sizes of educational interventions (see also Hattie, 2009). Our findings in relation to career aspirations are thus particularly worrisome. Students who might have the necessary set of skills to successfully pursue careers in science and technology but are schooled with other high-achieving students in selective school systems are most likely to suffer in their perceived

abilities and to be turned off from further postsecondary engagement in these fields.

What policy changes or intervention programs could be implemented to counter the negative effects of school-average ACH on ASC and consequently on career aspirations? One obvious way of countering the BFLPE would be the abolishment of all instances of selectivity and tracking in schools. However, even if it were possible to implement such wide-ranging school reforms, there would still be regional and local variations in achievement levels potentially leading to the BFLPE. Also, inclusion of learning-disadvantaged students into mainstream classes has been shown to negatively affect their ASC (Marsh, Tracey, & Craven, 2006), potentially offsetting positive effects of inclusion on the ASC of high-achieving students. Targeted intervention programs for countering the negative effects of the BFLPE on ASC and FUT would be an obvious alternative. Unfortunately, there is a shortage of research on such interventions (Marsh, Seaton, et al., 2008). Marsh and Craven (2002) speculated that reducing social comparison processes by de-emphasizing competition in schools and focusing more on assessment and feedback based on criterion reference might partially overcome the BFLPE. However, there is no direct evidence yet that employing these strategies would work and reduce the BFLPE. Lüdtke, Köller, Marsh, and Trautwein (2005), for example, found that teachers who emphasized an individualized frame-of-reference, highlighting individual progress rather than social comparisons, were able to influence the self-concept of their students positively but did not reduce the BFLPE.

In a similar vein, there is some evidence from meta-analyses of self-concept enhancement programs not specifically designed to counter the BFLPE (Haney & Durlak, 1998; O'Mara, Marsh, Craven, & Debus, 2006) that indicates that ASC could potentially be enhanced with appropriate interventions. This could be an especially effective way of countering the BFLPE on career aspirations, as the mediation models suggested that most of the effect was transmitted by individual ASC. Such self-concept interventions are particularly effective if they take into account the multidimensionality of ASC and focus on enhancing a single ASC dimension, such as mathematics or science (O'Mara et al., 2006). Successful interventions lend themselves to being implemented in school contexts: They typically include elements of (contingent) praise and feedback and some skills training (O'Mara et al., 2006) that could be used by schools trying to enhance ASC (Craven, Marsh, & Burnett, 2003; Craven, Marsh, & Debus, 1991). In fact, these strategies constitute good practices of instruction (Hattie, 2009) that would foster both ACH and ASC at the same time. Given the close relation between ASC and career aspirations at the individual student-level, they are also likely to exert a positive influence on students' plans for the future in the corresponding subjects, making big fish in big ponds aspire more.

References

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

Bachman, J. G. (1975). *Youth in transition: Data file documentation* (Vol. 2). Ann Arbor: University of Michigan, Institute for Social Research.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11,* 142–163. doi:10.1037/1082-989X.11.2.142

Betz, N. E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology, 28,* 399–410. doi:10.1037/0022-0167.28.5.399

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology, 98,* 550–558. doi:10.1037/a0018933

Butts, C. T. (2009). yacca: Yet Another Canonical Correlation Analysis Package (Version 1.1) [Computer software]. Retrieved from http://cran.r-project.org/web/packages/yacca/index.html

Byrne, B. M., & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology, 78,* 474–481.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among 15-year-olds in 34 countries. *Learning and Instruction, 20,* 2–17. doi:10.1016/j.learninstruc.2008.11.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Craven, R. G., Marsh, H. W., & Burnett, P. (2003). Cracking the self-concept enhancement conundrum: A call and blueprint for the next generation of self-concept enhancement research. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (Vol. 1, pp. 91–126). Greenwich, CT: Information Age.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology, 83,* 17–27. doi:10.1037/0022-0663.83.1.17

Craven, R. G., Marsh, H. W., & Print, M. (2000). Gifted, streamed and mixed-ability programs for gifted students: Impact on self-concept, motivation, and achievement. *Australian Journal of Education, 44,* 51–75.

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

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138. doi:10.1037/1082-989X.12.2.121

Ferry, T. R., Fouad, N. A., & Smith, P. L. (2000). The role of family context in a social cognitive model for career-related choice behavior: A math and science perspective. *Journal of Vocational Behavior, 57,* 348–364. doi:10.1006/jvbe.1999.1743

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology, 46,* 1454–1480. doi:10.1037/a0020204

Fouad, N. A. (2007). Work and vocational psychology: Theory, research, and applications. *Annual Review of Psychology, 58,* 543–564. doi:10.1146/annurev.psych.58.110405.085713

Fouad, N. A., & Smith, P. L. (1996). A test of a social cognitive model for middle school students: Math and science. *Journal of Counseling Psychology, 43,* 338–346. doi:10.1037/0022-0167.43.3.338

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58,* 479–514. doi:10.1146/annurev.psych.58.110405.085559

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53,* 455–467. doi:10.1007/BF02294400

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. doi:10.1146/annurev.psych.58.110405.085530

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A 10-year longitudinal study. *Self and Identity, 3,* 53–68. doi:10.1080/13576500342000040

Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career

development of women. *Journal of Vocational Behavior, 18,* 326–339. doi:10.1016/0001-8791(81)90019-1

Haney, P., & Durlak, J. A. (1998). Changing self-esteem in children and adolescents: A meta-analytic review. *Journal of Clinical Child Psychology, 27,* 423–433. doi:10.1207/s15374424jccp2704_6

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15,* 177–199. doi:10.1076/sesi.15.2.177.30432

Hattie, J. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37,* 449–481. doi:10.1016/S0883-0355(03)00015-6

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* London, United Kingdom: Routledge.

Hayes, A. F., & Preacher, K. J. (2010). Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. *Multivariate Behavioral Research, 45,* 627–660. doi:10.1080/00273171.2010.498290

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). New York, NY: McGraw-Hill.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

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

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15,* 309–334. doi:10.1037/a0020761

Jonsson, J. O., & Mood, C. (2008). Choice by contrast in Swedish schools: How peers' achievement affects educational choice. *Social Forces, 87,* 741–765. doi:10.1353/sof.0.0135

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Newbury Park, CA: Sage.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of *Culture's consequences*: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies, 37,* 285–320. doi:10.1057/palgrave.jibs.8400202

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1–21. doi:10.1207/s15327906mbr3001_1

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277. doi:10.1207/S15327906MBR3602_06

Lent, R. W., Brown, S. D., Brenner, B. R., Chopra, S. B., Davis, T., Talleyrand, R., & Suthakaran, V. (2001). The role of contextual supports and barriers in the choice of math/science educational options: A test of social cognitive hypotheses. *Journal of Counseling Psychology, 48,* 474–483. doi:10.1037/0022-0167.48.4.474

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45,* 79–122. doi:10.1006/jvbe.1994.1027

Lent, R. W., Brown, S. D., Sheu, H.-B., Schmidt, J., Brenner, B. R., Gloster, C. S., . . . Treistman, D. (2005). Social cognitive predictors of academic interests and goals in engineering: Utility for women and students at historically Black universities. *Journal of Counseling Psychology, 52,* 84–92. doi:10.1037/0022-0167.52.1.84

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., & Yeung, A. S. (2010, July). *The big fish little pond effect and ability-streaming: Effects of within-school ability grouping, moderators, and gender and ethnicity referents.* Paper presented at the 2010 International Congress of Applied Psychology, Melbourne, Australia.

Liu, W. C., Wang, C. K. J., & Parkins, E. J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology, 75,* 567–587. doi:10.1348/000709905X42239

Lopez, F. G., Lent, R. W., Brown, S. D., & Gore, P. A., Jr. (1997). Role of social-cognitive expectations in high school students' mathematics-related interest and performance. *Journal of Counseling Psychology, 44,* 44–52. doi:10.1037/0022-0167.44.1.44

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods, 16,* 444–467. doi:10.1037/a0024376

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

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. O. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13,* 203–229. doi:10.1037/a0012869

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34,* 120–131. doi:10.1016/j.cedpsych.2008.12.001

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337. doi:10.1023/A:1008929526011

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* New York, NY: Erlbaum.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39,* 99–128. doi:10.1207/s15327906mbr3901_4

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.* London, England: Academic Press.

Marsh, H. W. (1984). Self-concept, social comparison, and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal, 21,* 799–806.

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

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

Marsh, H. W. (1994). Using the National Educational Longitudinal Study of 1988 to evaluate theoretical models of self-concept: The Self-Description Questionnaire. *Journal of Educational Psychology, 86,* 439–456. doi:10.1037/0022-0663.86.3.439

Marsh, H. W. (2004). Negative effects of school-average achievement on academic self-concept: A comparison of the big-fish-little-pond effect across Australian states and territories. *Australian Journal of Education, 48,* 5–26.

Marsh, H. W. (2006). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology.* Leicester, United Kingdom: British Psychological Society.

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

Marsh, H. W., & Craven, R. G. (2002). The pivotal role of frames of reference in academic self-concept formation: The big-fish-little-pond effect. In F. Pajares & T. Urdan (Eds.), *Adolescence and education* (Vol. 2, pp. 83–123). Greenwich, CT: Information Age.

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

Marsh, H. W., & Hau, K. (2003). Big-fish-little-pond effect on academic self-concept.: A cross-cultural (26-country) test of the negative effects of

academically selective schools. *American Psychologist, 58,* 364–376. doi:10.1037/0003-066X.58.5.364

Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological–substantive synergies. *Contemporary Educational Psychology, 32,* 151–171. doi:10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. (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. doi:10.1207/s15327574ijt0604_1

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

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A., & Köller, O. (in press). Classroom climate and contextual effects: Methodological issues in the evaluation of group-level effects. *Educational Psychologist.*

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. O., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44,* 764–802. doi:10.1080/00273170903333665

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16,* 439–476. doi:10.1080/10705510903008220

Marsh, H. W., & O'Mara, A. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish-little-pond effect. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 24,* 51–72. doi:10.1024/1010-0652.a000004

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

Marsh, H. W., & Rowe, K. J. (1996). The negative effects of school-average ability on academic self-concept: An application of multilevel modelling. *Australian Journal of Education, 40,* 65–87.

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

Marsh, H. W., Tracey, D. K., & Craven, R. G. (2006). Multidimensional self-concept structure for preadolescents with mild intellectual disabilities: A hybrid multigroup–MIMC approach to factorial invariance and latent mean differences. *Educational and Psychological Measurement, 66,* 795–818. doi:10.1177/0013164405285910

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. doi:10.3102/0002831207306728

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

Marsh, H. W., & Yeung, A. S. (1997b). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89,* 41–54. doi:10.1037/0022-0663.89.1.41

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research, 46,* 816–841. doi:10.1080/00273171.2011.606716

McDonald, R. P. (1993). A general model for two-level data with responses missing at random. *Psychometrika, 58,* 575–585. doi:10.1007/BF02294828

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods & Research, 22,* 399–413. doi:10.1177/0049124194022003007

McFarland, C., & Buehler, R. (1995). Collective self-esteem as a moderator of the frog-pond effect in reactions to performance feedback. *Journal of Personality and Social Psychology, 68,* 1055–1070. doi:10.1037/0022-3514.68.6.1055

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equation modeling. *Psychological Methods, 10,* 259–284. doi:10.1037/1082-989X.10.3.259

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543. doi:10.1007/BF02294825

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56,* 177–196. doi:10.1007/BF02294457

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (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

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39,* 129–149. doi:10.1207/s15327906mbr3901_5

Mulkey, L. M., Catsambis, S., Steelman, L. C., & Crain, R. L. (2005). The long-term effects of ability grouping in mathematics: A national investigation. *Social Psychology of Education, 8,* 137–177. doi:10.1007/s11218-005-4014-6

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the United Kingdom, the UK countries and the world: The Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology, 31,* 629–656. doi:10.1080/01443410.2011.586416

National Center for Educational Statistics. (1986). *High School and Beyond, 1980: Sophomore cohort second follow-up (1984): Data file user's manual.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

O'Mara, A. J., Marsh, H. W., Craven, R. J., & Debus, R. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41,* 181–206. doi:10.1207/s15326985ep4103_4

Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: Results from the first OECD programme for international student assessment (PISA) 2000.* Paris, France: Author.

Organisation for Economic Co-operation and Development. (2005). *PISA 2003 technical report.* Paris, France: Author.

Organisation for Economic Co-operation and Development. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006.* Paris, France: Author.

Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Vol. 2. Data.* Paris, France: Author.

Organisation for Economic Co-operation and Development. (2009a). *PISA 2006 technical report.* Paris, France: Author.

Organisation for Economic Co-operation and Development. (2009b). *PISA data analysis manual* (2nd ed.). Paris, France: Author.

Pituch, K. A., & Stapleton, L. M. (2011). Hierarchical linear and structural equation modeling approaches to mediation analysis in randomized field experiments. In M. Williams & P. Vogt (Eds.), *The Sage handbook of*

innovation in social research methods (pp. 590–619). Thousand Oaks, CA: Sage.

Preacher, K. J. (2011). Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate Behavioral Research, 46,* 691–731. doi:10.1080/00273171.2011.589280

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers, 36,* 717–731. doi:10.3758/BF03206553

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16,* 93–115. doi:10.1037/a0022658

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18,* 161–182. doi:10.1080/10705511.2011.557329

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

Preckel, F., Götz, T., & Frenzel, A. (2010). Ability grouping of gifted students: Effects on academic self-concept and boredom. *British Journal of Educational Psychology, 80,* 451–472. doi:10.1348/000709909X480716

Pustjens, H., van de Gaer, E., Van Damme, J., & Onghena, P. (2004). Effect of secondary schools on academic choices and on success in higher education. *School Effectiveness and School Improvement, 15,* 281–311. doi:10.1080/09243450512331383222

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69,* 167–190. doi:10.1007/BF02295939

Rao, C. R. (1973). *Linear statistical inference and its applications.* London, England: Wiley.

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

Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling, 11,* 621–637. doi:10.1207/s15328007sem1104_7

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling, 16,* 583–601. doi:10.1080/10705510903203466

Savickas, M. L. (2002). Reinvigorating the study of careers. *Journal of Vocational Behavior, 61,* 381–385. doi:10.1006/jvbe.2002.1880

Savickas, M. L. (2005). The theory and practice of career construction. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 42–70). New York, NY: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* Boca Raton, FL: Chapman & Hall.

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

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond effect: Generalizability and moderation: Two sides of the same coin. *American Educational Research Journal, 47,* 390–433. doi:10.3102/0002831209350493

Seaton, M., Marsh, H. W., Dumas, F., Huguet, F., Monteil, J.-M., Régner, I., . . .Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology, 47,* 73–103. doi:10.1348/014466607X202309

Selig, J. P., & Preacher, K. J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Retrieved from http://quantpsy.org/medmc/medmc.htm

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

Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics, 35,* 26–53. doi:10.3102/1076998609345252

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* Boca Raton, FL: Chapman and Hall.

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

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13,* 290–312. doi:10.2307/270723

Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics, 33,* 230–251. doi:10.3102/1076998607307239

Spokane, A. R., Cruza-Guet, M. C., Brown, S. D., & Lent, R. W. (2005). Holland's theory of vocational personalities in work environments. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 24–41). New York, NY: Wiley.

Super, D. E. (1980). A life-span, life-space approach to career development. *Journal of Vocational Behavior, 16,* 282–298. doi:10.1016/0001-8791(80)90056-1

Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95,* 405–439. doi:10.1037/a0018938

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

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology, 101,* 853–866. doi:10.1037/a0016306

Tsui, A. S., Nifadkar, S. S., & Ou, A. Y. (2007). Cross-national, cross-cultural organizational behavior research: Advances, gaps, and recommendations. *Journal of Management, 33,* 426–478. doi:10.1177/0149206307300818

Tymms, P. (2001). A test of the big fish in a little pond hypothesis: An investigation into the feelings of 7-year-old pupils in school. *School Effectiveness and School Improvement, 12,* 161–181. doi:10.1076/sesi.12.2.161.3452

United Nations Development Program. (2009). *Human development index.* Retrieved October 20, 2009, from http://hdr.undp.org/en/statistics/data/

Valentine, J. C., & DuBois, D. L. (2005). Effects of self-beliefs on academic achievement and vice versa. In H. W. Marsh, R. G. Craven, & D. McInerney (Eds.), *The new frontiers of self research* (pp. 53–77). Greenwich, CT: Information Age.

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

VanderWeele, T. J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research, 38,* 515–544. doi:10.1177/0049124110366236

von Davier, M., Gonzalez, E., & Mislevy R. J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Vol. 2. Issues and methodologies in large scale assessments* (pp. 9–36). Princeton, NJ: IEA-ETS Research Institute.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and sub-*

stance abuse research (pp. 281–324). Washington, DC: American Psychological Association. doi:10.1037/10222-009

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25,* 68–81. doi:10.1006/ceps.1999.1015

Zeidner, M., & Schleyer, J. (1999). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology, 24,* 305–329. doi:10.1006/ceps.1998.0985

Received March 8, 2011
Revision received October 21, 2011
Accepted January 8, 2012 ■