Journal of Educational Psychology
2012, Vol. 104, No. 1, 87–107

0022-0663/11/$12.00 DOI: 10.1037/a0025734

The Internal/External Frame of Reference Model, Big-Fish-Little-Pond Effect, and Combined Model for Mathematics and Science

Mei-Shiu Chiu
National Chengchi University

The aim of this study was to integrate 2 theories on academic self-concept—that is, the internal/external frame of reference (I/E) model and the big-fish-little-pond effect (BFLPE)—into a combined model. The rationale for the combined model is that the I/E model actually focuses on internal/dimensional comparison (e.g., a comparison between one's own mathematics and science achievements) and the BFLPE focuses on external/social comparison (e.g., a comparison between one's own mathematics achievement and one's schoolmates'). The I/E, BFLPE, and combined models for mathematics and science were examined using multilevel analyses of 139,174 students in 4,231 schools in 27 countries. The results showed that the I/E model and the BFLPE fitted data well, and the combined model fitted data even better. The I/E model remained unchanged when the BFLPE was combined, suggesting distinctly differential roles of internal and external comparisons in constructing academic self-concepts. In addition, the combined model was examined for each country. The findings revealed that the predictions of the combined model were generally supported for data from most countries.

Keywords: academic self-concepts, achievement, the internal/external frame of reference model, the big-fish-little-pond effect

The understanding of psychological processes in constructing academic self-concepts is of paramount importance for educators. This understanding can facilitate suitable interventions provided by teachers who aim to help students obtain appropriate academic self-concepts and desirable learning outcomes, such as adequate allocation of time and effort, sense of control over learning, and maximization of achievements. These desirable outcomes in turn will benefit students in their pursuit of appropriate careers and sense of happiness (Marsh, Craven, & McInerney, 2003). As such, appropriate academic self-concepts have become one of the major objectives of education, and the understanding of the process in constructing academic self-concepts can bring both practical and theoretical benefits to education.

Two of the major theoretical models in relation to the psychological processes in constructing academic self-concepts to date are the internal/external frame of reference (I/E) model (Marsh, 1986) and the big-fish-little-pond effect (BFLPE; Marsh, 1987a). The basic I/E model highlights two major psychological processes (i.e., internal and external comparisons), while the BFLPE focuses on the process of external comparison only. Internal comparison refers to a psychological process in which individuals compare their own achievement in a particular domain (e.g., mathematics) with that in another domain (e.g., science). External comparison refers to a psychological process in which individuals compare their own achievement with group (e.g., school) achievements in a particular domain (e.g., mathematics) or in the academic domain. The overlap in external comparison posited by the two theoretical models suggests a possibility that there is a valid model that integrates the I/E model and the BFLPE.

The present study focuses on the domains of mathematics and science for two reasons. First, the two domains are two of the major school subjects in national curricula worldwide. By Grade 8, students are likely to have developed stable knowledge of their own and their classmates' or schoolmates' achievements in the two domains. This serves as a major source for student domain-specific self-concepts. Second, mathematics and science are different in their types of knowledge. Mathematics focuses on pattern and logic (Burton, 1994), which may be perceived by students as containing more procedural knowledge, while science appears to contain more declarative knowledge(e.g., earth science, life science, physics, chemistry, and environmental science), a system used by the Trends in International Mathematics and Science Study (TIMSS; International Association for the Evaluation of Educational Achievement [IEA], 2005). These two reasons suggest that mathematics and science are two different domains in terms of national curricula, knowledge types, and student perceptions, which provide a natural platform to investigate the I/E model and BFLPE, both of the theories emphasizing domain specificity, as is further addressed later.

To sum up, in theoretical terms, the major purpose of the present study was to propose and evaluate a combined model that integrates the I/E model and the BFLPE in the domains of mathematics and science. An evaluation of the combined model can advance our understanding of not only the I/E model and the BFLPE but

This article was published Online First October 3, 2011.

This research was supported by the National Science Council, Taiwan (NSC-95-2522-S-004-001; NSC 100-2511-S-004-004). The author wishes to express her gratitude to the editors for thoughtful suggestions and comments on the article.

Correspondence concerning this article should be addressed to Mei-Shiu Chiu, Department of Education, National Chengchi University, No. 64, Sec. 2, ZhiNan Road, Wenshan District, Taipei City 11605, Taiwan, R.O.C. E-mail: chium@nccu.edu.tw

also the concepts of internal and external comparisons theoretically and empirically. The theoretical backgrounds, theoretical rationales, and related studies of the three models are summarized in Table 1 and are explored in detail in the next sections.

The Internal/External Frame of Reference (I/E) Model

Marsh's (1986) I/E model is the most significant theory to date regarding both internal and external comparisons as major psychological processes in constructing academic self-concepts (Marsh, Craven, & Debus, 1999; Marsh & Hattie, 1996; Yeung & Lee, 1999). Marsh based his I/E model on three theoretical backgrounds, two major rationales, and the distinct domains of mathematics and verbal skills.

Theoretical Backgrounds of the I/E Model

The first background of the I/E model is Shavelson, Hubner, and Stanton's (1976) theory of hierarchical and multidimensional self-concepts. A higher order general self-concept can be classified into lower order nonacademic and academic self-concepts, each of which can be further classified into much lower order self-concepts. For example, academic self-concepts can be classified into self-concepts in all school domains (e.g., mathematics, science, history, and English). Shavelson et al.'s theory is supported by research results that show model fit using the statistical procedure of hierarchical confirmatory factor analysis and by research results that demonstrate that domain-specific self-concepts are more reliable in their relationship with important external criteria (e.g., academic achievement) than with general self-concept (Marsh, 1987b, Marsh, 1993; Marsh, Byrne, & Shavelson, 1988; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006; Marsh & Yeung, 1997b).

Marsh and Shavelson (1985) posited a revised Marsh/Shavelson model based on Shavelson et al.'s (1976) theory and the results of empirical studies showing that general self-concept comprises three lower order ones: mathematics, English, and general school (including other school subjects; Marsh, 1990b, Marsh, 1990c). Research on the Marsh/Shavelson model inevitably focuses on the two core domains of mathematics and verbal skills, which generates the consistent research finding that the correlation between mathematics and verbal self-concepts is much lower than that between mathematics and verbal achievements. This consistent finding—high correlations between achievements and low correlations between self-concepts, in non-matching domains—forms the second background of the I/E model (Marsh & Roche, 1996).

Social comparison theory is the third theoretical background for the basic I/E model in the aspect of external comparison. Individuals evaluate their own abilities by comparing them with others' abilities in a certain setting, in addition to objective, physical, and nonsocial standards for evaluation of their abilities (Festinger, 1954). When compared with outstanding others, individuals will lower their self-concepts, although interpersonal connectedness may slightly decrease this tendency (Cheng & Lam, 2007; Kemmelmeier & Oyserman, 2001) and individual inspiration may increase this tendency in selective settings (Burleson, Leach, Harrington, 2005).

Theoretical Rationales for the I/E Model

According to the I/E model, self-concept in any given domain is constructed by two comparison processes or frames of reference at the individual level. *Internal/ipsative-like/dimensional comparison or reference* across domains refers to a psychological process in which students compare their own achievements in one domain (e.g., mathematics) with their own in other domains (e.g., science). If a student's mathematics achievement is higher than science achievement, then this student's mathematics self-concept will be higher than this student's science self-concept even if the student's mathematics achievement is not good compared with other students'. The student's self-concept is constructed as, for example, "I am better at mathematics than science." *External/normative/social comparison* refers to a psychological process in which students compare their perceived achievement levels with other students' in a particular matching domain. The self-concept is constructed as, for example, "I am better at mathematics than most of my schoolmates."

Related Studies on the I/E Model

The I/E model is typically formulated as positive effects of achievements on self-concepts in matching domains (because of internal comparison) and negative effects of achievements on self-concepts in nonmatching domains (because of external comparison), along with controlling for correlations between achievements in different domains and between self-concepts in different domains. A typical example is related to the domains of mathematics and verbal skills: (a) The process of external comparison leads students with a high mathematics achievement to a high mathematics self-concept and leads students with a high verbal achievement to a high verbal self-concept and (b) the process of internal comparison leads students with a high mathematics achievement to a low verbal self-concept and leads students with a high verbal achievement to a low mathematics self-concept. The framework for the I/E model can be depicted using an interactively controlled path system (see Panels A and B in Figure 1).

A significant number of studies have provided evidence supporting the I/E model across cultures (Abu-Hilal & Bahri, 2000; Marsh, 1986, Marsh, 1989; Marsh & Köller, 2004; Mui, Yeung, Low, & Jin, 2000; Plucker & Stocking, 2001; Skaalvik & Rankin, 1995, Skaalvik & Rankin, 2002). Most studies on the I/E model focus on the two distinctly different domains of mathematics and verbal skills (e.g., Goetz, Frenzel, Hall, & Pekrun, 2008; Marsh, 1989; Marsh & Hau, 2004; Skaalvik & Rankin, 1992, Skaalvik & Rankin, 1995). Marsh and Hau (2004) provided evidence for a cross-cultural generalizability of the I/E model in a comparison of mathematics and verbal skills across 26 countries. The results of Möller, Pohlmann, Köller, and Marsh's (2009) meta-analysis of 69 data sets from 63 empirical studies by 2006 focusing on the mathematics and verbal skills also provides strong support for the predictions of the basic I/E model across countries, gender groups, and age groups.

Later studies have gradually and successfully extended to other distinct domains (i.e., native vs. nonnative languages; e.g., Marsh, Kong, & Hau, 2001). Bong's (1998) study also focused on the domains of mathematics and verbal skills, but the mathematics domains included algebra, geometry, and chemistry, while the

Table 1
A Comparison of the I/E Model, the BFLPE, and the Combined Models

Variable	I/E model	BFLPE	I/E model and BFLPE combined
Theoretical backgrounds	1. Multidimensional academic self-concepts 2. High correlations between achievements and low correlations between self-concepts in nonmatching domains 3. Social comparison theory	1. Multidimensional academic self-concept 2. Social comparison theory	1. Multidimensional academic self-concepts 2. High correlations between achievements and low correlations between self-concepts in nonmatching domains 3. Social comparison in matching domains 4. Strong, reciprocal, and positive relations between achievements and self-concepts in matching domains at the individual level
Theoretical rationales	1. Internal/ipsative-like/dimensional comparison in nonmatching domains at the individual level 2. External/normative/social comparison in matching domains at the individual level	Social comparison in matching domains at the group level	1. Internal comparison across domains at the individual level 2. External comparison in matching domains at the group level
Examples of model operation in past studies	(1) Student mathematics achievement can positively predict student mathematics self-concept but negatively predict student science self-concept and (2) student science achievement can positively predict student science self-concept but negatively predict student mathematics self-concept, controlling for the correlation between mathematics and science achievements and that between mathematics and science self-concepts (Chiu, 2008).	School mathematics achievements can negatively predict student mathematics self-concept, controlling for the positive effect of student mathematics achievement (Marsh et al., 2007).	
Hypotheses (models) in the present study	Hypothesis 1 (Model 2) Hypothesis 2 (Model 10)	Hypothesis 3 (Model 4) Hypothesis 4 (Model 12)	Hypothesis 5 (Model 7) Hypothesis 6 (Model 15)

Note. I/E = internal/external frame of reference; BFLPE = big-fish-little-pond effect.

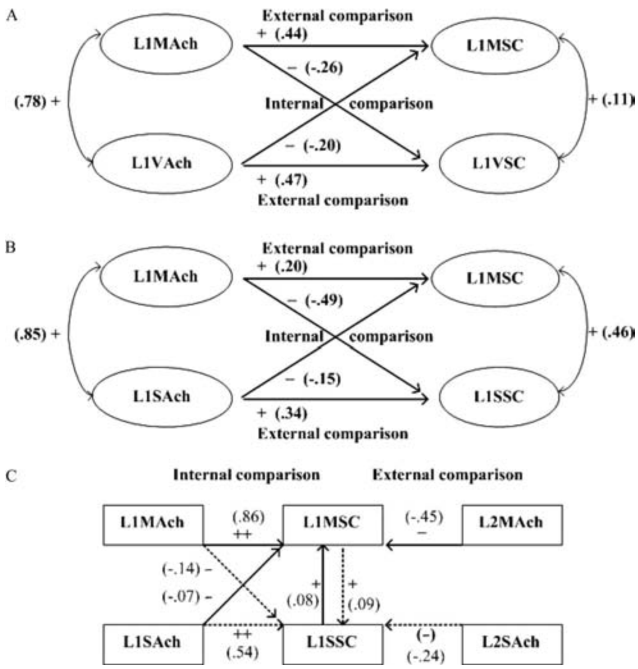

Figure 1. Comparison of the results of an examination of the internal/external frame of reference (I/E) model—used in the Marsh and Hau (2004) study of mathematics and verbal skills (Panel A) and in the Chiu (2008) study of mathematics and science (Panel B)—with the results of an examination of the model combining the I/E model and the big-fish-little-pond effect in the present study (Panel C). The numbers in parentheses show the major results presented in the three studies. A path or link with "+" represents a positive prediction (or relation), and that with "–" represents a negative one. In Panel C, arrows with solid lines represent the function structure used in Model 7 (see Table 2), and arrows with dashed lines represent that used in Model 15 (see Table 3). The "(–)" in Panel C indicates a present finding that has not been examined in past studies. L1 = Level 1 (student level); L2 = Level 2 (school level); MAch = mathematics achievement; MSC = mathematics self-concept; VAch = verbal achievement; VSC = verbal self-concept; SAch = science achievement; SSC = science self-concept.

verbal domains consisted of English, Spanish, and American history. The inclusion of diverse domains to the constructs of mathematics and verbal skills may partly account for the results of the study, which failed to support the predictions of the I/E model. Marsh and Yeung (2001) reanalyzed the data of Bong's study and found that the constructs of achievements and self-concepts in Spanish were much different from those in all other domains. The factor correlations between the global self-concept construct of the mathematics skills and its specific ones of algebra, geometry, and chemistry were .79, .69, and .51, respectively (p. 401), which revealed that chemistry was worth a study that regarded it as a different construct from mathematics.

Research has also found that the I/E model is supported for the domains of mathematics and science even if the correlations between mathematics and science self-concepts are slightly larger than those between mathematics and verbal self-concepts across cultures. Chiu (2008) applied structural equation modeling (SEM), the same statistical procedure used by Marsh and Hau (2004), to examine the predictions of the basic I/E model for the domains of mathematics and science. The parameters obtained for the critical paths predicted by the I/E model successfully supported the model's predictions despite the slightly large correlation between mathematics and science self-concepts (cf. Panels A and B in Figure 1). The results suggest that the generalization of the I/E model has been successfully extended from the traditional domains of mathematics and verbal skills to the domains of mathematics and science. Möller, Streblow, Pohlmann, and Köller's (2006) study, however, showed a different story. They included two verbal domains (German and English) and two numerical domains (mathematics and physics) in a single model. The results showed that German and English self-concepts were positively predicted by German and English achievements and negatively predicted by mathematics and science achievements, while mathematics and science self-concepts were positively predicted by mathematics and science achievements and negatively predicted by German and English achievements. Three reasons for the inconsistent findings regarding the predictions of the I/E model may be possible. First, the predictions of the I/E model are supported only for distinctly different domains if related and unrelated domains are simultaneously considered in a single model. German/English and mathematics/science were too related to significantly show the contrast effect of achievements on self-concepts in different domains or the psychological process of internal comparison. If there are only two related domains in a single model, as in Chiu's study, the results are likely to fit the prediction of the I/E model. Second, the models in Möller et al.'s study did not include the procedure of controlling for relations between achievements and between self-concepts in the four domains. This controlling procedure is especially necessary for related domains, as there are usually high correlations between achievements and between self-concepts in related domains. The predictions of the I/E model are likely to become salient if this controlling procedure is included. Third, differences between cultures, test items, and research participants are also possible reasons.

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

Theoretical Backgrounds of the BFLPE

The major theoretical backgrounds of the BFLPE include multidimensional academic self-concept and social comparison theory. Like the I/E model, the BFLPE was developed on the basis of Shavelson et al.'s (1976) theory of multidimensional self-concepts. Unlike the studies of the I/E model, which have depicted relations between achievements and self-concepts across different domains, studies of the BFLPE have generally focused on academic self-concept, either in general or in a particular domain.

The second background of the BFLPE is social comparison theory, as described in the I/E model. Students have a tendency to compare their achievement with their peers' achievements on the same campus regardless of college reputations. Davis (1966) used a frog-pond metaphor for the campus and proposed that students in selective colleges were likely to experience relative deprivation in career aspirations. Marsh (1987a; Marsh & Parker, 1984) further introduced the frame-of-reference hypothesis based on research on psychophysics, created the term *BFLPE*, and extended the research participants to school-level students. There are also some minor theoretical underpinnings from social psychology and sociology (e.g., social judgment and relative deprivation; Marsh & Hau, 2003).

Theoretical Rationales for the BFLPE

The BFLPE regards external comparison or reference as a major psychological process in constructing academic self-concepts. Two processes are included in the BFLPE: Students (a) compare their achievements with others' in their settings (e.g., classes and schools) and (b) use the relative impression of their abilities to construct part of their academic self-concept (Marsh, 1991). The BFLPE indicates that equally able students will have lower self-concepts if they are in a high-achieving group and have higher self-concepts if they are in a low-achieving group. Although there are slightly positive effects of group-average achievement on student self-concept (i.e., reflected glory or assimilation effects), the negative effect of group-average achievements on student self-concept (i.e., contrast effects) is much stronger than the positive one, and the net group effects are generally negative (Marsh, 1987a; Marsh, Kong, & Hau, 2000). As a result, BFLPE researchers have suggested the disadvantage of selective schools or gifted classes because able students are likely to have a low academic self-concept and career aspiration in these selective contexts. They also stress the advantage of full-time special education classes or schools because disadvantaged students are likely to suffer from low self-concepts when placed in regular classes or schools (Marsh & Hau, 2003; Preckel, Zeidner, Goetz, & Schleyer, 2008).

Related Studies on the BFLPE

The BFLPE is typically formulated as a negative effect of group (school/class) average achievement, controlling for a positive effect of student achievement, on student academic self-concept in general or in a particular domain. A significant number of studies have provided evidence that supports the BFLPE across cultures (e.g., Marsh, Chessor, Craven, & Roche, 1995; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007; Zeidner & Schleyer, 1999). Marsh and Hau (2003) further provided positive support for a cross-cultural generalizability of the BFLPE across 26 countries. The BFLPE is supported by not only correlational studies but also experimental ones across academic self-concepts in general, mathematics, and reading (e.g., McFarland & Buehler, 1995). Some studies find a BFLPE at the class level (e.g., Liu, Wang, & Parkins, 2005; Manger & Eikeland, 1997; Rogers, Smith, & Coleman, 1978; Tymms, 2001; Wong & Watkins, 2001), and others find it at the school level (e.g., Marsh, 1984, Marsh, 2004; Marsh et al., 2000, Marsh et al., 2007). The BFLPE is robust even when the BFLPE model is incorporated with diverse constructs (e.g., school/student socioeconomic status; Marsh, 1987a). Studies on the BFLPE typically focus on one single school domain, and the most popular domains are still the two core school subjects: mathematics (e.g., Manger & Eikeland, 1997; Marsh et al., 2000, Marsh et al., 2007) and verbal skills, such as English or reading (e.g., Marsh, 1987a; Marsh & Rowe, 1996; Rogers et al., 1978). An exception is Marsh and Hau's study, in which mathematics, science, and verbal achievement scores were averaged for each student, and the domain is the global academic achievement. Domain specificity is still an important characteristic of the BFLPE.

The Combined Model Integrating the I/E Model and the BFLPE

A combined model that integrates the I/E model and the BFLPE is likely to be a valid structure in that the I/E model and the BFLPE share similar theoretical backgrounds, although they have appeared to be two separate lines of studies to date. The fact that the I/E model and the BFLPE have the same theoretical background of social comparison theory but different rationales and operations for external comparison suggests a valid combined model, which operates the conceptions of external and internal comparisons distinctively in a single model and therefore may advance the understanding of the basic I/E and BFLPE models.

Theoretical Backgrounds of the Combined Model

The combined model has four backgrounds. The first three backgrounds are the same as those for the I/E model and the BFLPE: (a) multidimensional academic self-concept, (b) high correlations between achievements and low correlations between self-concepts in the different domains, and (c) social comparison theory. The fourth theoretical background is the strong, reciprocal, and positive relations between achievements and self-concepts in matching domains at the individual level, as indicated by three lines of research. First, the relationships between ability-related beliefs and achievements are regarded as results of personal matters by a number of theories on motivation—for example, expectancy-value theory (Wigfield & Eccles, 2000) and self-efficacy (Zimmerman, 1995). Second, Festinger (1954) ascertained that nonsocial sources readily available for self-evaluation (e.g., test results and homework assessments) might occur before social comparison. Students may focus on concerns about their own achievements based on some relatively objective standards and form some self-images of abilities before they compare their own achievements with other students' in their contexts, which further elaborates their self-images. The third line of research focuses on the reciprocal-effects models (REMs) of the relationship between achievements and self-concepts based on a developmental perspective using longitudinal data. Results of REM studies show that achievement and self-concept are both a cause and an effect for each other substantially and dynamically (Guay, Marsh, & Boivin, 2003; Helmke & Van Aken, 1995; Kurtz-Costes & Schneider, 1994; Marsh, Chanal, & Sarrazin, 2006; Marsh & Craven, 2006; Marsh, Hau, & Kong, 2002; Marsh & O'Mara, 2008; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Marsh & Yeung, 1997a; Skaalvik & Hagtvet, 1990) with few exceptions (Hoge, Smit, & Crist, 1995; Ma & Xu, 2004; Muijs, 1997; Pottebaum, Keith, & Ehly, 1986). The REM is especially supported under the condition of domain specificity (Valentine, DuBois, & Cooper, 2004).

As such, further consideration may be needed if one is to completely attribute the strong link between achievement and self-concept to social comparison when both variables are constructs of personal phenomena. One therefore can further speculate that the predictions of the basic I/E model (see Panels A and B in Figure 1) may be explained as a result entirely based on the psychological process of internal comparison (see the left side of Panel C in Figure 1) when other salient and relevant variables of external comparison are incorporated into the same model (see the right side of Panel C in Figure 1). While the I/E model may need a salient external variable to supplement the notion of external comparison, the BFLPE appears to be a theory significantly and basically based on the frame-of-reference hypothesis at the external level (Marsh & Parker, 1984) or social comparison theory (Marsh, 2004; Marsh & Hau, 2003). These theoretical backgrounds make the BFLPE a suitable theory to be incorporated into the I/E model.

Theoretical Rationales for the Combined Model

In the combined model, the psychological processes that construct self-concept in a particular domain include both internal and external comparisons, the same terms as those used in the basic I/E model. However, the operation of external comparison is different from that in the basic I/E model. The combined model maintains the notion that the entire operation of the basic I/E model (e.g., see Panels A and B in Figure 1) is based on internal comparison and the operation of external comparison can be supplemented by the operation in the BFLPE, which forms another process in constructing academic self-concept. Internal and external comparisons work together for students to construct their multidimensional academic self-concepts.

Internal comparison functions on the basis of individuals' awareness of their relative strengths in different domains. Students regard their own achievements in different domains as objective, physical, and nonsocial standards at the individual level. The association effect of achievements on self-concepts occurs in matching domains, and the contrast effect of achievements on self-concepts occurs in nonmatching domains. For example, students' awareness of their relative abilities in mathematics and verbal skills is based on comparison between their mathematics and verbal achievements at the same level. Internal comparison relies on comparison between objective, physical, and nonsocial results of tests in different domains. There are positive effects of mathematics/verbal achievements (the objective standard) on mathematics/verbal self-concepts because students associate their achievements with their self-concepts in matching domains. There are negative effects of mathematics/verbal achievements on verbal/mathematics self-concepts because students contrast their own achievements in different domains and slightly revise self-concepts in different domains. Internal comparison is operated as a comparison made within students between their achievements across domains, including matching and nonmatching ones; in other words, the entire operation of Marsh's (1986) I/E model (e.g., see Panels A and B in Figure 1) captures the full essence of the psychological process of internal comparison.

External comparison functions on the basis of individuals' drive to compare with others if objective, physical, and nonsocial standards are not available (Suls, 1977). External comparison can be conceptualized as individual students' making a comparison between themselves and a generalized other based on group (class/school) average achievement, as described in the BFLPE. There are also association and contrast effects in the BFLPE, but association effects (e.g., higher school status leading to higher self-concepts) are always weaker than contrast effects (e.g., higher school-average achievements leading to lower self-concepts; Marsh, Kong, & Hau, 2000). In addition, external comparison is domain-specific. As such, external comparison can be operated as a net negative effect of group-average achievements on student self-concepts in a particular domain.

Related Studies on the Combined Model

Studies on the I/E model support a positive external-comparison effect of student achievement on student self-concepts in matching domains, while studies on the BFLPE find a negative external-comparison effect of class/school achievements on student self-concepts, also in matching domains. A question that emerges is, What is the net effect of the "positive" external-comparison effect as indicated in the basic I/E model (Panels A and B in Figure 1) and the "negative" external-comparison effect as posited in the BFLPE? One way to answer this question is to compare the results obtained from examinations of the I/E model, the BFLPE, and the combined model. If one finds that the effects indicated by the I/E model remain unchanged because of the inclusion of the BFLPE, then one can infer that the psychological processes underlying the two theories are likely to be distinctly different. In other words, the positive path from student achievements to student self-concepts may actually be based on the psychological process of internal comparison (different from the explanation of the basic I/E model; cf. Panels A and B and the left side of Panel C in Figure 1), while the negative path from group achievements to student self-concepts may result from the psychological process of external comparison, as explained by the BFLPE (cf. the right side of Panel C in Figure 1).

To date, there are two sets of studies partially aiming to examine a combined model that integrates the I/E model and the BFLPE. The first set of studies was conducted by Marsh (1990a, 1994), who placed student/school mathematics/reading (English) achievements and student mathematics/reading (English) self-concepts together into a single model using SEM. The results of his studies support both the basic I/E and BFLPE models, with student achievements positively predicting student self-concepts in matching domains, student achievements negatively predicting student self-concepts in nonmatching domains, and school achievements negatively predicting student self-concepts in matching domains. The two studies by Marsh, however, were operated at the student level, with school-level achievements being disaggregated for each student for the use of SEM, a typical statistical procedure for examining the I/E model by past studies. Further, the theoretical rationales for Marsh's combined model is not clearly addressed, especially for the rationales for the same conceptions but different operations of external comparison in the basic I/E and BFLPE models. Finally, the effects of school achievements on student self-concepts in nonmatching domains were investigated in Marsh (1990a) but not in Marsh (1994). Marsh (1990a) found small positive effects of school achievements on student self-concepts in nonmatching domains. He attributed the results to a rationale similar to that of the basic I/E model: The psychological process of internal comparison creates this counterbalancing effect that large negative effects of group achievements on individual self-concepts in matching domains should accompany small positive effects of group achievements on individual self-concepts in nonmatching domains. Theoretically, the BFLPE is based on the psychological process of external comparison, by which individuals engage in assessing their relative positions within the local group. The process of internal comparison between group achievements in different domains posited by Marsh (1990a), on the other hand, implies that individuals assess relative strengths of group achievements in different domains, which are based on data obtained from comparison with other groups within a larger context. This process may attract individuals to group concerns, rather than personal concerns, and may further attract individuals to the reflected-glory effect, which indicates that students in selective schools will have high self-concepts and will reduce the BFLPE (Marsh et al., 2000). For example, if students in a school consider that their school's mathematics achievement is lower than the other schools' and that their school's science achievement is higher than the other schools' together, it may be hard for the students themselves to infer that they are good at mathematics because of their bad school achievement in mathematics. As such, the counterbalancing effect due to the process of internal comparison between group achievements in different domains may be weak. Practically and technically, the extremely high correlations (normally above .90) between group achievements in different domains for most empirical data may reduce the possibility of the existence of the process of internal comparison between group achievements and may dictate caution in the explanation of research results because of the problem of multicollinearity in regression analysis. (As a result, controlling for correlations between groups' achievements in different domains is a necessary procedure and was used by Marsh, 1990a, in his SEM model.)

The second set of studies is Möller and Köller's (2001) experimental approach to studying the I/E model. The operation of external comparison in their study was to provide participants with feedback about their performances as above, below, or equal to average compared with those of generalized others' on the same figure task. The operation of internal/dimensional comparison was to provide participants with information about their performances as better than, worse than, or equal to their performances on the word task. In line with the predictions of the I/E model, participants provided with information about their performances on the figure task as above the average performance on the figure task or as better than their own performances on the word task had an increased self-concept on the figure task, and vice versa. Although Möller and Köller's study focused on the I/E model, their operation of external comparison was similar to that in the BFLPE, because social comparison information is directly operated as comparison with generalized others in a particular domain in their experiments. The typical operation of external comparison in the basic I/E model, on the other hand, is achievements leading to self-concepts in matching domains, the data of which, however, is still at the individual level; the conception of social comparison is not saliently operated. The inclusion of the basic BFLPE in the basic I/E model is likely to further clarify the conception of external comparison.

Statistical Advances

In addition to theoretical concerns, there is a need to determine a proper methodology to examine the combined model. Past studies normally examined the predictions of the I/E model using SEM (e.g., Marsh & Hau, 2004) or multiple regression (e.g., Williams & Montgomery, 1995). The basic I/E model is generally operated using the statistical procedure of SEM. Two typical studies that successfully modeled and supported the predications of the basic I/E model are Marsh and Hau's (2004) study focusing on the domains of mathematics and verbal skills and Chiu's (2008) study focusing on those of mathematics and science (see Panels A and B in Figure 1). A note to make is that the statistical procedure of SEM that formulates the basic I/E model emphasizes the juxtaposition of all four constructs (e.g., mathematics/science achievement/self-concept) in a single model. The correlation between mathematics and science achievements and that between mathematics and science self-concepts are set as controlling in the same model, in addition to the four major predicting paths from mathematics/science achievements to mathematics/science self-concepts. As such, it appears to be important to include the four constructs in the same model and to include the controlling variables in addition to the major predictors when one aims to model the basic I/E model using regression, as is used in the present study. A note to make is that one can set correlations between two variables as controlling in SEM; one can also place likely confounding/influential factors as predictors into a regression formula as controlling in (multilevel) regression analysis.

On the other hand, the BFLPE was examined using analysis of variance (ANOVA; e.g., Manger & Eikeland, 1997; Rogers et al., 1978), path analysis (e.g., Marsh, 1987a; Marsh & Parker, 1984), SEM (e.g., Marsh, 1994), or multilevel analysis (e.g., Marsh & Rowe, 1996). Multilevel analysis appears to be a recently available technique that has replaced ANOVA and path analysis as the standard tool for examining the BFLPE, in which group achievement and student achievement serve as two predictors to predict student self-concept in a particular domain, as is used in the present study.

SEM is a powerful tool to model the structural relations between variables at a single level, especially at the individual level. Although SEM is gradually being developed to include the analysis of multilevel data (Du Toit & Du Toit, 2001; Muthén & Muthén, 2006; Stapleton, 2002), the technique is still at a preliminary stage. Multilevel analysis allows researchers to specify and test complex multi- and cross-level models, by including both individual- and group-level variables altogether, which addresses the issues of both internal and external comparisons altogether. The present study, therefore, chooses multilevel analysis as the major statistical method to examine the posited hypotheses.

One may ask: What is the best model for formulating the combined model? The earlier review of literature on typical and advanced statistical procedures for modeling the basic I/E model and the BFLPE suggests useful criteria for the best model to formulate the combined model by using statistical procedures as consistent as possible with the procedures for the two basic models. Given the condition that multilevel regression analysis is used, two criteria emerge from the review of the literature. First, include all four constructs at the student level (i.e., mathematics/science achievement/self-concept) in the same model; in other words, predict a self-concept in a domain with all of the other three variables, when one aims to formulate the basic I/E model. Second, include one matching-domain achievement at the student level and one at the group level to predict a self-concept in a particular domain, when one aims to formulate the basic BFLPE. To combine the two criteria, the best model for the combined model will be, for example: student mathematics self-concept (the dependent variable for both the basic I/E model and the BFLPE) predicted by student mathematics achievement (a major predictor for the I/E model and a controlling variable for the BFLPE), science achievement (a major predictor for the I/E model), and science self-concept (a controlling variable for the I/E model) and by school mathematics self-concept (a major predictor for the BFLPE). This formulation is represented in Model 7 of Table 2 with student mathematics self-concept as the dependent variable and in Model 15 of Table 3 with student science self-concept as the dependent variable.

The Present Study

In statistical terms, the major purpose of the present study was to investigate the combined model that integrated the I/E model and the BFLPE with student mathematics and science self-concepts as dependent variables or outcomes and mathematics and science achievements at the student and school levels as independent variables or predictors using multilevel (regression) analysis. The I/E model and the BFLPE were also examined before the investigation of the combined model because there were few studies focusing on the juxtaposition of mathematics and science for the I/E model and no studies to date focusing on science for the BFLPE. Despite this, it is reasonable to expect that the basic I/E model and the BFLPE are supported using multilevel analysis. The reason for this expectation is that Chiu's (2008) study has successfully extended the I/E model to the domains of mathematics and science using SEM, even though an innovation of the present study was to use multilevel analysis to operate the I/E model. The predictions of the BFLPE are also supported, since multilevel analysis is a typical and advanced procedure for examining the theory. Further, if we operated only the group-level variable for a single domain, the opportunity for successful support for the BEFLPE would increase, because past studies normally focused on a single domain, even though science is rarely a domain for examining the BFLPE.

Six hypotheses were posited and organized into three sets. The purpose of Hypotheses 1–2 was to examine the basic I/E model at the student level, that of Hypotheses 3–4 was to examine the basic BFLPE model at the school level, and that of Hypotheses 5–6 was to examine the combined model at both student and school levels using three-level multilevel models but without incorporating country-level aggregates. The results of the examinations of these hypotheses are shown by the models presented in Tables 2–3. The results of Models 7 and 15 are also presented in Panel C of Figure 1.

> *Hypothesis 1:* Student mathematics self-concept is positively predicted by student mathematics achievement but negatively predicted by student science achievement, controlling for student science self-concept (see Model 2 in Table 2).

> *Hypothesis 2:* Student science self-concept is positively predicted by student science achievement but negatively predicted by student mathematics achievement, controlling for student mathematics self-concept (see Model 10 in Table 3).

> *Hypothesis 3:* Student mathematics self-concept is negatively predicted by school mathematics achievement, controlling for student mathematics achievement (see Model 4 in Table 2).

Table 2
Models of Student Mathematics Self-Concept as a Function of Mathematics and Science Achievements for Students, Schools, and Countries and Student Science Self-Concept

Model description	Model 1 (Empty model)	Model 2 (I/E 1)		Model 3 (I/E 2)		Model 4 (BFLPE 1)		Model 5 (BFLPE 2)		Model 6 (BFLPE 3)		Model 7 (I/E 1 + BFLPE 1)		Model 8 (I/E 1 + BFLPE 1 controlling for L3MAch)	
		Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*
Fixed effect															
L1MAch		.81***	.09	.81***	.09	.80***	.06	.79***	.06	.80***	.06	.86***	.09	.86***	.09
L1SAch		−.17***	.04	−.12*	.05							−.14***	.04	−.14***	.04
L1SSC		.09**	.03									.08*	.03	.08*	.03
L2MAch						−.50***	.05			−.20**	.08	−.45***	.05	−.45***	.05
L2SAch								−.48***	.06	−.31***	.09				
L3MAch														−.25**	.08
Random effect															
L1	.88	.73		.74		.74		.74		.74		.73		.73	
L2	.06	.09		.09		.05		.05		.05		.05		.05	
L3	.06	.36		.42		.11		.16		.13		.10		.05	
Deviance	383,688.09	359,494.12		360,667.79		358,857.60		358,759.63		358,702.72		357,601.47		297,580.97	
No. of estimated parameters	4	7		6		6		6		7		8		9	

Note. I/E = internal/external frame of reference model; BFLPE = big-fish-little-pond effect; Est = parameter estimate; L1 = Level 1 (student level); L2 = Level 2 (school level); L3 = Level 3 (country level); MAch = mathematics achievement; SAch = science achievement; SSC = science self-concept.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3
Models of Student Science Self-Concept as a Function of Mathematics and Science Achievements for Students, Schools, and Countries and Student Mathematics Self-Concept

Model description	Model 9 (Empty model)	Model 10 (I/E 1)		Model 11 (I/E 2)		Model 12 (BFLPE 1)		Model 13 (BFLPE 2)		Model 14 (BFLPE 3)		Model 15 (I/E 1 + BFLPE 1)		Model 16 (I/E 1 + BFLPE 1 controlling for L3SAch)	
		Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*
Fixed effect															
L1MAch		−.09*	.05	−.02	.06							−.07	.05	−.07	.05
L1SAch		.52***	.06	.51***	.06	.54***	.04	.54***	.04	.54***	.04	.54***	.06	.54***	.06
L1MSC		.09**	.04									.09*	.04	.09*	.04
L2MAch								−.30***	.04	−.19*	.09				
L2SAch						−.28***	.04			−.11	.08	−.24***	.03	−.24***	.04
L3MAch														−.24**	.07
Random effect															
L1	.85	.77		.77		.77		.77		.77		.76		.76	
L2	.08	.09		.09		.07		.07		.07		.07		.07	
L3	.08	.23		.29		.13		.12		.12		.11		.05	
Deviance	378,656.65	365,229.86		366,410.83		365,722.82		365,698.21		365,683.26		364,715.22		364,694.73	
No. of estimated parameters	4	7		6		6		6		7		8		9	

Note. I/E = internal/external frame of reference model; BFLPE = big-fish-little-pond effect; Est = parameter estimates; L1 = Level 1 (student level); L2 = Level 2 (school level); L3 = Level 3 (country level); MAch = mathematics achievement; SAch = science achievement; MSC = mathematics self-concept.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Hypothesis 4: Student science self-concept is negatively predicted by school science achievement, controlling for student science achievement (see Model 12 in Table 3).

Hypothesis 5: Student mathematics self-concept is positively predicted by student mathematics achievement but negatively predicted by student science achievement and school mathematics achievement, controlling for student science self-concept (see Model 7 in Table 2).

Hypothesis 6: Student science self-concept is positively predicted by student science achievement but negatively predicted by student mathematics achievement and school science achievement, controlling for student mathematics self-concept (see Model 15 in Table 3).

Two sets of auxiliary analyses were used to present additional support for the posited combined models (Hypotheses 5–6; i.e., see Models 7 and 15 in Tables 2–3) based on empirical data: (a) The hypotheses in relation to the basic I/E model (Hypotheses 1–2) were examined without controlling for the effect of self-concepts in nonmatching domains (Models 3 and 11), and (b) the hypotheses in relation to the basic BFLPE (Hypotheses 3–4) were examined with the inclusion of group self-concepts in nonmatching domains but without group self-concepts in matching domains (Models 5 and 13) and with the inclusion of group self-concepts in both matching and nonmatching domains (Models 6 and 14).

Furthermore, the generalizability of the combined models (see Models 7 and 15 in Tables 2–3) was tested using two-level multilevel analyses for each of the 27 countries (see Table 4). Following this logic, the combined models (Models 7 and 15) were examined using three-level multilevel models controlling for the effect of country achievement (see Models 8 and 16 in Tables 2–3), which took into account one of the major factors of country differences and might result in much better combined models for the total population.

Method

Data Source and Sample

The data analyzed in the present study were taken from the database of the TIMSS of 2003, which was compiled by the IEA (2005). The TIMSS 2003 database included data on mathematics and science achievements of Grade 8 students from 47 countries. While the 47 countries supplied data on student mathematics self-concepts, only 28 countries had data on student science self-concepts. Among the 28 countries, the Syrian Arab Republic was the only one without data on country mathematics and science achievements (as presented in the TIMSS international mathematics and science reports, which are described in detail in the next section), so further analysis did not include the data from this country. Hence, the final sample was 139,174 students in 4,231 schools from the remaining 27 countries, which had all the necessary data for the present analysis. The 27 countries are presented in Table 4.

Measures and Data Preparation

The present study used data from the student, school, and country levels, which was consistent with the three-level sampling

Table 4
Selected Parameter Estimates Obtained by Multilevel Analysis Based on the I/E and BFLPE Combined Models per Separate Country

			Model 7 (predicting L1MSC)									Model 15 (predicting L1SSC)								
			L1MAch		L1SAch		L1SSC		L2MAch			L1MAch		L1SAch		L1MSC		L2SAch		
Country	Country code	*N*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Fit	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Fit
Total		139,174	.86	.09	–.14	.04	.08	.03	–.45	.05	FS	*–.07*	.05	.54	.06	.09.	.04	–.24	.03	PS
Australia	36	4,791	.74	.03	–.10	.02	.14	.02	–.29	.04	FS	–.25	.04	.50	.04	.17	.02	*–.06*	.04	PS
Bahrain	48	4,199	.47	.02	–.09	.02	.12	.02	–.07	.03	FS	*.03*	.03	.26	.02	.13	.02	–.16	.03	PS
Botswana	72	5,150	.38	.02	–.18	.02	.12	.02	–.05	.02	FS	*.02*	.02	.29	.02	.12	.02	–.11	.03	PS
Chile	152	6,377	.65	.03	*–.01*	.03	.03	.01	–.47	.03	PS	*–.03*	.03	.29	.02	.03	.01	–.13	.03	PS
Taiwan	158	5,379	.64	.03	–.09	.02	.21	.02	–.20	.04	FS	–.21	.04	.41	.05	.26	.02	–.06	.03	FS
Palestinian Nat'l Auth	275	5,357	.41	.03	–.08	.03	.18	.02	–.11	.03	FS	*.07*	.03	.28	.03	.18	.03	*–.05*	.03	PS
Ghana	288	5,100	.23	.03	*.02*	.02	.14	.02	–.07	.03	PS	*.05*	.03	.35	.03	.14	.02	*–.03*	.03	PS
Hong Kong	344	4,972	.74	.03	–.08	.02	.14	.02	–.40	.03	FS	–.27	.03	.55	.02	.15	.02	–.20	.03	FS
Iran	364	4,942	.48	.02	*–.01*	.02	.12	.02	–.54	.06	PS	.09	.02	.24	.02	.13	.02	–.34	.07	NS
Israel	376	4,318	.49	.03	*–.01*	.02	.05	.02	–.16	.04	PS	*–.02*	.03	.43	.03	.05	.02	–.15	.03	PS
Italy	380	4,278	.66	.02	*–.03*	.02	.08	.02	–.30	.02	PS	–.07	.03	.33	.03	.10	.03	*–.05*	.03	PS
Japan	392	4,856	.63	.02	–.19	.02	.20	.02	–.15	.02	FS	–.21	.03	.46	.03	.23	.02	–.09	.04	FS
Jordan	400	4,489	.41	.03	*–.01*	.03	.14	.02	–.12	.03	PS	.07	.03	.25	.03	.14	.02	–.07	.03	NS
Korea, South	410	5,309	.60	.03	–.09	.03	.15	.01	–.08	.02	FS	–.17	.03	.44	.02	.18	.02	*.03*	.02	PS
Malaysia	458	5,314	.69	.03	–.12	.03	.27	.02	–.27	.03	FS	–.42	.03	.47	.03	.31	.02	*–.03*	.03	PS
Morocco	504	2,943	.23	.05	*.06*	.04	.21	.03	–.10	.03	PS	*.04*	.04	.23	.03	.21	.03	–.09	.04	PS
New Zealand	554	3,801	.72	.02	–.07	.03	.09	.02	–.35	.03	FS	–.17	.03	.54	.02	.10	.02	–.16	.03	FS
Norway	578	4,133	.63	.02	–.07	.02	.14	.02	*–.06*	.04	PS	–.21	.04	.39	.02	.19	.03	*.02*	.03	PS
Philippines	608	6,917	.49	.04	–.14	.03	.13	.02	–.28	.04	FS	–.12	.03	.43	.03	.13	.02	–.10	.04	FS
Saudi Arabia	682	4,295	.38	.02	*.00*	.03	.12	.03	–.15	.03	PS	.06	.02	.28	.02	.12	.03	*–.05*	.03	NS
Singapore	702	6,018	.85	.02	–.50	.02	.13	.01	–.11	.02	FS	–.61	.03	.76	.02	.14	.01	–.06	.03	FS
South Africa	710	8,952	.59	.03	–.10	.03	.08	.02	–.42	.03	FS	.15	.04	.34	.03	.09	.02	–.28	.03	NS
Tunisia	788	4,931	.46	.02	*–.03*	.02	*–.01*	.01	–.11	.02	PS	–.05	.02	.30	.02	*–.01*	.02	–.11	.02	FS
Egypt	818	7,095	.35	.03	*–.01*	.03	.15	.02	*–.04*	.03	PS	*–.03*	.03	.40	.04	.15	.02	–.07	.03	PS
United States	840	8,912	.69	.03	–.18	.03	*–.01*	.02	–.21	.02	FS	–.11	.04	.46	.04	*.00*	.03	–.16	.03	FS
England	926	2,830	.65	.03	–.21	.04	.08	.03	–.24	.04	FS	–.30	.05	.65	.03	.08	.03	–.16	.04	FS
Scotland	927	3,516	.56	.03	–.10	.03	.15	.02	–.23	.04	FS	*–.01*	.03	.45	.03	.15	.02	–.10	.03	PS
M			.55	.03	–.09	.03	.12	.02	–.21	.03		–.10	.03	.40	.03	.14	.02	–.10	.03	
SD			.16	.01	.10	.01	.06	.01	.14	.01		.17	.01	.13	.01	.07	.01	.08	.01	

Note. Separate multilevel analysis was conducted for each country based on the I/E and BFLPE combined models (i.e., Models 7 and 15 in Tables 2 and 3, respectively), in which the analyses were conducted for the total population (N = 139,174). The italicized estimates are not significant at p = .05. I/E = internal/external frame of reference model; BFLPE = big-fish-little-pond effect; L1 = Level 1 (student level); L2 = Level 2 (school level); MSC = mathematics self-concept; SSC = science self-concept; MAch = mathematics achievement; SAch = science achievement; Est = parameter estimates; FS/PS/NS = full/partial/no support for the predictions of the combined models (Models 7 and 15); Nat'l Auth = National Authority.

structure of the TIMSS. Country-level (Level 3) measures included country mathematics and science achievements as shown next.

Country mathematics achievement (labeled L3MAch here). The values of country mathematics achievement were taken from the Average Scale Score column at Exhibit 1.1 in the TIMSS 2003 International Mathematics Report (Mullis, Martin, Gonzalez, & Chrostowski, 2004, p. 34). The mean and the standard error of the 27 countries' mathematics achievements were 452.41 and 17.87, respectively.

Country science achievement (L3SAch). The 27 values of country science achievement were taken from Exhibit 1.1 in the TIMSS 2003 International Science Report (Martin, Mullis, Gonzalez, & Chrostowski, 2004, p. 36) with $M = 463.04$ and $SE = 16.95$.

These country achievement scores were computed as "the average of the weighted means for each of the five plausible values" (Gonzalez, Galia, Arora, Erberber, & Diaconu, 2004, p. 296). In the TIMSS, five plausible values were obtained as the authentic estimates of mathematics ability (i.e., TIMSS variables bsmmat01–bsmmat05) and five values as those of science ability (i.e., TIMSS variables bsssci01–bsssci05) for each student based on the scaling procedure of item response theory. The present study applied similar procedures for obtaining the country achievement scores to the estimation of achievements at the student and school levels in order to make the data at the three levels more directly comparable (the procedures are described in *Step 1* in the next paragraph). The procedures for dealing with missing data, facilitating multilevel analysis, and avoiding collinearity or multicollinearity were also used. These procedures consisted of the following five steps, which partly influenced the definitions of the student-level and school-level measures.

Step 1. The two means of the five plausible values, estimated on the basis of item response theory and multiple imputation (IEA, 2005, pp. 2–51), in relation to mathematics and science achievements for each student were computed. The country achievement scores described in the TIMSS reports, also described earlier, can be obtained using the means of the five plausible values, but not any one of the single plausible values, by activating either student-level weight. Therefore, the use of the means of the five plausible values at the student level appears to match the use of the country achievements better than the use of either one plausible value at the student level.

Step 2. The problem of missing data had to be solved here as the proportions of missing data were 3.4% for mathematics self-concept and 2.9% for science self-concept. To deal with missing data, the procedure of multiple imputation (Olinsky, Chen, & Harlow, 2003; Schafer, 1997) was implemented by Monte Carlo Markov chain algorithm using the software of LISREL 8.72 (Du Toit & Du Toit, 2001; Jöreskog & Sörbom, 2001, Jöreskog & Sörbom, 2005). Five new data sets were generated by the procedure using the data of the four measures at the student levels and the other seven auxiliary variables taken from the same TIMSS database. The new data sets remained the same as the original one except for the estimated values that replaced the missing data. Multilevel analysis was conducted on each of the five data sets, the results produced by each analysis were combined using a formula suggested by Rubin (1987, pp. 76–77; Schafer & Graham, 2002), and the figures presented in Tables 2–4 are the combined results. The definitions of the four student-level measures were partly influenced by Steps 1–2, as shown next.

Student mathematics achievement (L1MAch, based on TIMSS variables bsmmat01–bsmmat05). There were five plausible values of mathematics achievement for each student provided by the TIMSS database. The score was the mean of the five plausible values in relation to mathematics achievement for each student ($M = 450.14$, $SE = 0.33$). The types of knowledge tested on the mathematics achievement scale included the subdomains of data ($M = 455.97$, $SE = 0.31$), fractions/numbers ($M = 451.29$, $SE = 0.33$), geometry ($M = 448.37$, $SE = 0.33$), and measurement ($M = 450.11$, $SE = 0.32$).

Student science achievement (L1SAch, based on TIMSS variables bsssci01–bsssci05). There were also five plausible values of science achievement for each student. The score was the mean of the five plausible values in relation to science achievement for each student ($M = 458.20$, $SE = 0.33$). The types of knowledge tested on the science achievement scale included the subdomains of earth science ($M = 456.41$, $SE = 0.32$), life science ($M = 460.67$, $SE = 0.32$), physics ($M = 455.49$, $SE = 0.33$), chemistry ($M = 456.23$, $SE = 0.31$), and environmental science ($M = 465.48$, $SE = 0.31$).

Student mathematics self-concept (L1MSC, based on TIMSS-derived variable bsdmscl). This measure was the variable of Index of Students' Self-Confidence in Learning Mathematics, with data directly provided by the TIMSS database. Detailed descriptions of the way by which the measure was obtained can be found in Section 1-10 of Supplement 3 in the *TIMSS 2003 User Guide for the International Database* (IEA, 2005) and is summarized as follows. The scores were computed on the basis of students' responses to four items on a 4-point Likert scale ranging from 1 (*Agree a lot*) to 4 (*Disagree a lot*). The four items were "I usually do well in mathematics," "Mathematics is more difficult for me than for many of my classmates" (reversed-scored), "Mathematics is not one of my strengths," and "I learn things quickly in mathematics." The means of students' responses to the four items were calculated and categorized as follows: 1 = high (mean $\leqq 2$), 2 = medium ($2 <$ mean < 3), and 3 = low (mean $\geqq 3$), a variable provided in the TIMSS database. The categories were reverse-coded in the present study so that larger numbers could represent higher self-concepts and vice versa (i.e., 3 = high self-concept; 2 = medium; 1 = low; $M = 2.20$, $SE = 0.00$). The reverse coding is necessary because the models posited in the present study need the coding of the variables to represent their meanings in a positive direction (i.e., the higher the score, the higher the self-report of self-concept). This can facilitate the interpretation of the parameter estimates obtained from the multilevel analysis (cf. Figure 1 and Tables 2–4).

Student science self-concept (L1SSC, based on TIMSS-derived variable bsdsscl). This measure was the variable of Index of Students' Self-Confidence in Learning Science, with data directly provided by the TIMSS database. Detailed descriptions of the way by which the measure was obtained can be found in Section 1-11 of Supplement 3 in the *TIMSS 2003 User Guide* (IEA, 2005). The scores were calculated on the basis of students' responses to four items, which were "I usually do well in science," "Science is more difficult for me than for many of my classmates (reversed-scored)," "Science is not one of my strengths," and "I learn things quickly in science." The methods of scaling, coding, and scoring for student science self-concept were the same as those for student mathematics self-concept. The scores were reverse-coded in the present study so that a larger number of the value could represent a higher self-concept and vice versa ($M = 2.35$, $SE = 0.00$).

The reliability and validity of these four student-level measures can be found in Chiu's (2008) study, which has tested the basic I/E model using SEM for mathematics and science and using the same database as that used in the present study. The internal consistency reliabilities (Cronbach's alpha) for the four measures were .97 (mathematics achievement), .97 (science achievement), .70 (mathematics self-concept), and .69 (science self-concept) for the total world sample. Detailed descriptions of the reliability coefficients for each of the countries can also be found in Table 3 of Chiu. For construct validity, the result of a confirmatory factor analysis revealed a clear structure of the four factors with desirable values of fit indices (p. 241 in Chiu, 2008). The factor loadings for the items of both mathematics and science achievements were all above .90, and those of mathematics and science self-concepts were above .70 for the positively worded items but low for negatively worded ones (.29 ~ .47; cf. Table 1 of Chiu, 2008).

Step 3. The achievement data at the school level were calculated as the average of the weighted means for each of the five plausible values for each school, a procedure that is the same as that used for obtaining the data for country achievement, as described in the beginning of the Measures and Data Preparation section. As a result, the school-level measures were defined as follows:

School mathematics achievement (L2MAch). The score was the average of the five plausible values in relation to mathematics achievement for the students within a school.

School science achievement (L2SAch). The score was the average of the five plausible values in relation to science achievement for the students within a school.

Step 4. All student, school, and country measures were transformed into standardized z scores ($M = 0$, $SD = 1$). This procedure can facilitate the presentation and interpretation of the results from multilevel analysis and reduce the potential problem of multicollinearity (Aiken & West, 1991, p. 35) without being at the expense of a reliable and valid estimation of parameters. The procedure was especially necessary for the present data as the measures were quite different in their score ranges (e.g., 1–3 for student self-concepts and 264–605 for country mathematics achievements). A similar procedure was successfully used by Marsh et al. (2000, 2007) and some studies reported by Raudenbush and Bryk (2002). With the procedure, the effect estimates presented in Tables 2–4 can be explained as similar to standardized regression weights in traditional regression analyses.

Step 5. The degrees of the problem of multicollinearity among the student measures were also worth an examination, as we needed to include three student-level predictors when examining the regression models in relation to the I/E model and include two group-level predictors when examining the models in relation to the BFLPE. The correlations between the four student-level measures were quite small (see Table 5), except for the correlation between student mathematics and science achievements (.88). The correlation between country mathematics and science achievements was .96 ($p < .001$), and the correlation between school mathematics and science achievements was also .96 ($p < .001$). According to Hair, Black, Babin, Anderson, and Tatham (2006), the presence of high correlations (.90 and larger) between predictors in multiple regression analysis is a sign of multicollinearity. One of the direct indicators of multicollinearity is variance inflation factor (VIF), in which a value larger than 10 is viewed as an indication of a high degree of multicollinearity. When we formulated a regression equation using L1MSC as the dependent variable and the other three student-level measures as the predictors, the values of the VIFs of the three predictors were 4.67 for L1MAch, 4.71 for L1SAch, and 1.04 for L1SSC, which means that the standard errors of the regression coefficients of the three predictors increased by 2.16 ($= \sqrt{4.67}$), 2.17 ($= \sqrt{4.71}$), and 1.02 ($= \sqrt{1.04}$) times, respectively, because of multicollinearity. The results of analysis by applying the country- and school-level measures, respectively, to the same equation revealed that all the VIF values of the achievements were above 10, with 18.32 for L3MAch, 14.82 for L3SAch, 13.41 for L2MAch, and 12.91 for L2SAch. These results revealed that the problem of multicollinearity was not serious at the student level but was serious at the group levels. As such, it appears sensible not to include both mathematics and science achievements at the same group level (e.g., L2MAch and L2SAch) as predictors in a regression model because of the problem of multicollinearity, for which any regression estimates obtained for each predictor may be unstable and need to be explained with caution.

Table 5
Correlations Between Student-Level Measures

Variable	MAch	SAch	MSC
SAch	.88***		
MSC	.16***	.10***	
SSC	.01**	.08***	.22***

Note. MAch = mathematics achievement; SAch = science achievement; MSC = mathematics self-concept; SSC = science self-concept.
** $p < .01$. *** $p < .001$.

Statistical Analysis

A series of multilevel analyses at the three levels of students, schools, and countries was performed to explore the utility of the predictors. Multilevel analysis was suitable for the present purpose because the students were nested within schools that were grouped into countries. Typical linear regression analysis can only deal with data from a single level (e.g., either the student level or the country level); if one aggregates data at heterogeneous levels to either level, the variance estimates will be biased. Multilevel analysis can separate variances derived from different levels of data and take account of different sample sizes in each group. Multilevel analysis, therefore, can generate more unbiased estimates of fixed effects and standard errors than can the traditional analysis of linear regression.

The measures at the three levels were placed into the procedure of multilevel analyses using the software of HLM 6.02 (Raudenbush, Bryk, & Congdon, 2005). Weights were activated for student and school levels, as suggested by the TIMSS (IEA, 2005). The predictors were centered around their grand means of individuals, schools, and countries, so that the incremental effect from groups (e.g., the BFLPE) could be directly identified (Lüdtke et al., 2008;

Raudenbush & Bryk, 2002). The analysis was limited to random-intercept models, as the focus of the present study was on direct effects of the measures and no interaction effects were explored. Detailed procedures for multilevel analysis can also be found in Raudenbush, Bryk, Cheong, Congdon, and Du Toit (2004) and Snijders and Bosker (1999).

To illustrate the procedures used in the present study, we first considered a one-way random-effect ANOVA model, or empty model, using student mathematics self-concepts as the dependent variable (or outcome); no predictor was added to the model. Equation 1 shows the empty student, or Level 1, model:

$$L1MSC_{ijk} = \pi_{0jk} + e_{ijk}, \tag{1}$$

in which L1MSC_{ijk} is the mathematics self-concept of student i in school j and country k, π_{0jk} (intercept) is the average self-concepts of students in each of the schools in each country, and e_{ijk} is the Level 1 error term. The e_{ijk} is assumed to be normally distributed, with a mean equal to zero and a constant Level 1 variance.

In the school, or Level 2, model, shown in Equation 2,

$$\pi_{0jk} = \beta_{00k} + \gamma_{0jk}, \tag{2}$$

where π_{0jk} is the average student achievements in each of the schools in each country, β_{00k} is the average school achievements in each of the countries, and γ_{0jk} is the Level 2 error term. The γ_{0jk} is assumed to be normally distributed, with a mean equal to zero and a constant Level 2 variance.

In the country, or Level 3, model, seen in Equation 3,

$$\beta_{00k} = \gamma_{000} + u_{00k}, \tag{3}$$

where β_{00k} is the average school achievements in each of the countries, or the country achievements; γ_{000} is the average country achievement; and u_{00k} is the Level 3 error term. The u_{00k} is assumed to be normally distributed, with a mean equal to zero and a constant Level 3 variance.

The empty model just shown can function as a typical ANOVA test. For the present case, we knew the proportions of the total variance of L1MAch (i.e., the outcome) contributed by countries, schools, and students. This is an unconditional model, with no predictors. When we added effective predictors to the ANOVA model, we could see changes in the Level 1, Level 2, and Level 3 residual variances.

Second, we considered a model with one student predictor, one school predictor, and one country predictor. In the Level 1 model, seen in Equation 4,

$$L1MSC_{ijk} = \pi_{0jk} + \pi_{1jk} \times L1MAch_{ijk} + e_{ijk}, \tag{4}$$

where student mathematics self-concepts L1MSC were predicted by student mathematics achievements L1MAch, with intercept π_{0jk}, slope π_{1jk}, and residual term e_{ijk}.

There were two Level 2, or school-level, models for the π_{0jk} and π_{1jk} in Equation 4. The intercept model for π_{0jk} included school mathematics achievement (L2MAch) as the predictor and was allowed to vary across schools, by setting its residual term (γ_{0jk}) to be estimated, as seen in Equation 5:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} \times L2MAch_{jk} + \gamma_{0jk}. \tag{5}$$

The inclusion of L2MAch in the intercept model of Equation 5 could be used to investigate the simple effect of school achievements on student achievements (β_{01k}). The slope model for the π_{1jk} in Equation 4 did not include any predictor and was not allowed to vary across schools, as seen in Equation 6:

$$\pi_{1jk} = \beta_{10k}. \tag{6}$$

The first Level 3 model (see Equation 7), that is, the intercept model for β_{00k} (the average school mathematics achievements in each of the countries), was predicted by country mathematics achievement (L3MAch), which was used to investigate the direct effect of L3MAch on L1MSC. The intercept (β_{00k}) was allowed to vary across countries, and its residual term (u_{00k}) was allowed to be estimated. The other two Level 3 models (see Equations 8–9) are slope models for β_{01k} and β_{10k}, respectively, which did not include any predictors and were not allowed to vary across countries; in other words, their residual terms were not estimated:

$$\beta_{00k} = \gamma_{000} + \gamma_{001} \times L3MAch_k + u_{00k}, \tag{7}$$

$$\beta_{01k} = \gamma_{010}, \tag{8}$$

and

$$\beta_{10k} = \gamma_{100}. \tag{9}$$

A mixed model was yielded by substituting the π_{0jk} and π_{1jk} in Equation 4 with the two Level 2 models (Equations 5–6) and by substituting the β_{00k}, β_{01k}, and β_{10k} in Equations 5–6 with the three Level 3 models (Equations 7–9). Each term in the mixed model should be explained as functioning under the condition of the mixed model, in which L1MAch, L2MAch, and L3MAch were added as predictors. Given the condition of the mixed model, γ_{000} is the grand mean of the intercepts across countries; γ_{001}, γ_{010}, and γ_{100} are the estimates representing the effects of L3MAch, L2MAch, and L1MAch, respectively.

Multilevel analysis allowed for a number of Level 1, Level 2, and Level 3 predictors to be freely included in the models, and each model was regarded as a whole system. For example, the effect of L1MAch (γ_{100}) was significant in predicting L1MSC given a certain mixed model. For each of the fixed effects (i.e., γ_{000}, γ_{001}, γ_{010}, and γ_{100}), multilevel analysis produced the estimates of a regression coefficient and a standard error, which could be used to determine the significance of the coefficient. Random effects were indicated by residual terms (i.e., e_{ijk}, γ_{0jk}, and u_{00k}), which showed the residual variances in the Level 1 model (the variance of e_{ijk}), Level 2 models (the variance of γ_{0jk}), and Level 3 models (the variance of u_{00k}). Tables 2–3 report the estimates of the fixed effects for the predictors. The residual variance components of the intercepts for three levels are presented separately for each model so that one can judge to what extent residual variance components are influenced by different predictors being included in the models. The deviance of a model serves a purpose similar to that of residual variance components of a model. The difference in deviance between two nested models has a chi-square distribution, and the degree of freedom is the difference in the numbers of parameters estimated between two nested models. The hypothesis that a more general model is a better fit than a simpler one can be supported when the result of the chi-square difference test is significant. Hox (2002) recommended the chi-square difference test for examining the improvement of model fit in multilevel analysis.

Results

Two sets of multilevel regression analyses were implemented. Student mathematics self-concept (see Models 1–8 in Table 2) and student science self-concept (see Models 9–16 in Table 3) served as dependent variables. The first model in each set of analyses was an empty model. Student, school, and country measures were gradually added to the remaining models. The two sets of analyses were presented in a similar manner to facilitate discussion.

The Variance Sources of Student Self-Concepts

Models 1 and 9 were empty models. The values of random effects (variance components) indicated that 88% of the total variance of student mathematics self-concept was within students, 6% ($p < .001$) was between schools, and 6% ($p < .001$) was between countries (Model 1). In addition, 85% of the total variance of student science self-concept was within students, 8% ($p < .001$) was between schools, and 8% ($p < .001$) was between countries (Model 9). The random effects of the schools and of the countries were all significant, which meant that there were significant differences between schools and between countries in regard to student mathematics and science self-concepts.

The proportions of the variance sources of self-concepts from schools and countries were significant, but the magnitudes were not large (all below 10%). The significant results, however, suggested that the effects could not be ignored and that we needed to control school and country influences by including these variables in the multilevel models (cf. Tables 2–3). In addition, one may need to let each country tell its own story by examining the posited models for each country (cf. Table 4) and by delving into the context of each country and the school contexts within each country, if possible, in future research.

The I/E Model

The results of Models 2 and 10 supported the predictions of the I/E model (Hypotheses 1–2). Student mathematics self-concept was positively predicted by student mathematics achievement (.81) but negatively predicted by student science achievement (–.17), controlling for student science self-concept (.09; Hypothesis 1; see Model 2 in Table 2). Student science self-concept was positively predicted by student science achievement (.52) but negatively predicted by student mathematics achievement (–.09), controlling for mathematics self-concept (.09; Hypothesis 2; see Model 10 in Table 3). Compared with the results of Models 2 and 10, the results of Models 3 and 11 showed that the negative effect of student science achievement on student mathematics self-concept became less strong (–.12) and that of student mathematics achievement on student science self-concept became nonsignificant (–.02) without controlling for self-concepts in nonmatching domains. The results imply that Models 2 and 10 can more accurately grasp the essence of the theory of the basic I/E model as formulated by the typical statistical procedure for examining it using SEM (in which correlations between the achievements in two domains are controlled for, as indicated by Panels A and B in Figure 1) than can Models 3 and 11. Therefore, Models 2 and 10, rather than Models 3 and 11, were used as part of the combined model.

We found that the residual variances in student self-concepts changed significantly from .88 (Model 1) to .73 (Model 2) and from .85 (Model 9) to .77 (Model 10), but those in Levels 2 and 3 increased from .06 (Level 2) and .06 (Level 3) in Model 1 to .09 (Level 2) and .36 (Level 3) in Model 2 and from .08 (Level 2) and .08 (Level 3) in Model 9 to .09 (Level 2) and .23 (Level 3) in Model 10. The results imply that the I/E model explains student self-concepts at the student level but that there remains a significant amount of unexplained variance at the group level. The results of chi-square difference tests in deviance between the two nested models (i.e., Model 1 vs. Model 2; Model 9 vs. Model 10) revealed that Models 2 and 10 fitted the data better than did Models 1 and 9 (Models 1 vs. 2: $\chi^2 = 24{,}193.97$ [383,688.09 − 359,494.12], $df = 3$ [7 − 4], $p < .001$; Models 9 vs. 10: $\chi^2 = 13{,}426.79$ [378,656.65 − 365,229.86], $df = 3$ [7 − 4], $p < .001$).

The BFLPE

The BFLPE was supported at the school level for both mathematics and science self-concepts as dependent variables (Hypotheses 3–4). Table 2 shows that student mathematics self-concept was negatively predicted by school mathematics achievement (–.50), controlling for student mathematics achievement (.80; Model 4; Hypothesis 3). Student mathematics self-concept was also negatively predicted by school science achievement (–.48), controlling for student mathematics achievement (.79; Model 5). Further, student mathematics self-concept was negatively predicted by both school mathematics achievement (–.20) and school science achievement (–.31), controlling for student mathematics achievement (.80; Model 6). The residual variance components at Level 2 were the same (.05) across the three models (Models 4–6), which implied that the effects of school mathematics and science achievements were similar. Therefore, it appeared to be redundant and harmful if we included effects of both school mathematics and science achievement in the same BFLPE model, given the problem of multicollinearity. In choosing a better operation for the BFLPE between Models 4 and 5, Model 4 (using school mathematics achievement) appeared to be better than Model 5 (using school science achievement) when student mathematics self-concept was the dependent variable based on the theoretical rationale of domain specificity for the BFLPE.

Table 3 shows that student science self-concept was negatively predicted by school science achievement (–.28), controlling for student science achievement (.54; Model 12; Hypothesis 4). Student science self-concept was also negatively predicted by school mathematics achievement (–.30), controlling for student science achievement (.54; Model 13). Student science self-concept was negatively predicted by school mathematics achievement (–.19) but not significantly by school science achievement (–.11), controlling for student science achievement (.54; Model 14). Model 12 was considered to be the best among Models 12–14 when student science self-concept was the dependent variable based on the same rationale for choosing the best BFLPE model among Models 4–6 that included the concern for the problem of multicollinearity and the theoretical rationale for the BFLPE.

The Combined Model Integrating the I/E Model and the BFLPE

The I/E model and BFLPE were combined for the cases of student mathematics and science self-concepts as dependent variables. In Model 7 of Table 2, student mathematics self-concept was positively predicted by student mathematics achievement (.86) and negatively predicted by student science achievement (–.14) and school mathematics achievement (–.45), controlling for student science self-concept (.08), which supported Hypothesis 5. Model 7 (the combined model) was better than both Model 2 (the basic I/E model; χ^2 = 1,892.65 [359,494.12 – 357,601.47], *df* = 1 [8 – 7], $p < .001$) and Model 4 (the basic BFLPE; χ^2 = 1,256.13 [358,857.60 – 357,601.47], *df* = 2 [8 – 6], $p < .001$). In Model 15 of Table 3, student science self-concept was positively predicted by student science achievement (.54) and negatively predicted by student mathematics achievement (–.07, though not significant) and school science achievement (–.24), controlling for student mathematics self-concept (.09). The results supported Hypothesis 6. Model 15 (the combined model) was better than both Model 10 (the basic I/E model; χ^2 = 514.64 [365,229.86 – 364,715.22], *df* = 1 [8 – 7], $p < .001$) and Model 12 (the basic BFLPE; χ^2 = 1,007.60 [365,722.82 – 364,715.22], *df* = 2 [8 – 6], $p < .001$). Panel C in Figure 1 summarizes the results of Models 7 and 15.

These results show strong relations between self-concepts and achievements for the matching domains and relatively weak relations for the nonmatching domains, which are consistent with the findings of related studies (e.g., Marsh & Hau, 2004). The results suggest that matching-domain effects are stronger than nonmatching-domain effects. Nonmatching-domain effects, however, cannot be ignored, as they are significant and are in a negative direction, while matching-domain effects are in a positive direction.

The Combined Model Across 27 Countries

Country-to-country variations in the key parameter estimates for the combined model (Models 7 and 15) were explored by making the key parameters random at the country level before examining the generalizability of the combined model for the 27 countries. The results showed that, for Model 7, there were small but significant random effects associated with country for student mathematics achievement (.07), student science achievement (.02), and school mathematics achievements (.02; all *p*s $< .001$); for Model 15, there were also small but significant random effects associated with country for student mathematics achievement (.04), student science achievement (.05), and school science achievements (.01; all *p*s $< .001$). The results showed that there was significant country variation in the degree of fit to the combined model and that there appeared to be a need to examine the generalizability of the combined model across countries.

The generalizability of the combined models formulated as Models 7 and 15 (see Tables 2–3) was tested for each of the 27 countries. The results presented in Table 4 show that the data from these countries generally supported the prediction of the combined model as formulated by Model 7 and partially supported that as formulated by Model 15. The results were consistent with those of Models 7 and 15 for the total population (see Tables 2–3). In predicting student mathematics self-concept, the mean effects of student mathematics achievement (.55), student science achievement (–.09), and school mathematics achievements (–.21) across the 27 countries were consistent with the directions of the respective parameter estimates of Model 7 for the total population (see Table 2). In predicting student science self-concept, the mean effects of student mathematics achievement (–.10), student science achievement (.40), and school science achievements (–.10) across the 27 countries were consistent with the directions of the respective parameter estimates of Model 15 for the total population (see Table 2).

To describe the findings in more detail, we grouped the results of the countries into three categories in terms of their degrees of support for Hypotheses 5 and 6: (a) Full support: All predictions with directions in the combined model are completely supported. That is, student mathematics (science) self-concept is significantly positively predicted by student mathematics (science) achievement and significantly negatively predicted by student science (mathematics) achievement and school mathematics (science) achievement. (b) No support: At least one prediction is in the opposite direction. That is, student mathematics (science) self-concept is significantly negatively predicted by student mathematics (science) achievement or significantly positively predicted by either student science (mathematics) achievement or school mathematics (science) achievement. (c) Partial support: At least one prediction is not significant, except for the results that belong to the no support category. That is, student mathematics (science) self-concept is not significantly predicted by student mathematics (science) achievement, student science (mathematics) achievement, or school mathematics (science) achievement. The results presented in Table 4 show that the combined model formulated as Model 7 was fully supported by data from 16 countries and partially supported by data from 11 countries. The combined model formulated as Model 15 was fully supported by data from nine countries, partially supported by data from 15 countries, and not supported by data from four countries.

Given the variation across countries in their degrees of fit to the combined model, it is reasonable to test whether the inclusion of country achievements, one of the significant country-level variables, as predictors in the basic combined models (Models 7 and 15) would reduce the variation. The results presented in Model 8 in Table 2 show that in predicting student mathematics self-concept, the regression estimates of the basic combined model (Model 7) remained unchanged, controlling for the effect of country mathematics achievement, which was significant (–.25). The random effect at the country level decreased from .10 (Model 7) to .05 (Model 8). Model 8 (the combined model, controlling for country achievement) appeared to be better than Model 7 (the combined model; χ^2 = 60,020.50 [357,601.47 – 297,580.97], *df* = 1 [9 – 8], $p < .001$). Model 16 in Table 3 shows similar results. In predicting student science self-concept, the regression estimates of the basic combined model (Model 15) remained unchanged, controlling for the effect of country science achievement (–.24, significant). The random effect at the country level decreased from .11 (Model 15) to .05 (Model 16). Model 16 (the combined model, controlling for country achievement) appeared to be better than Model 15 (the combined model; χ^2 = 20.49 [364,715.22 – 364,694.73], *df* = 1 [9 – 8], $p < .001$).

The inclusion of the random effect of countries in the models allowed the variation of countries to be considered, although we are not sure what particular effects occurring at the country level actually contributed to the unexplained variance. The inclusion of country achievements in the models (Models 8 and 16) reduced the unexplained variance. The country achievements, however, may only be a reflection of multiple contextual factors in the country (e.g., curricula, pedagogies, and academic values); the negative country effects are difficult to explain. As such, Models 7 and 15 remain the best candidates of the combined models in the present study.

Discussion

The (I/E + BFLPE) Combined Model for Roles of Internal and External Comparisons in Constructing Academic Self-Concepts

The I/E model and the BFLPE, respectively, were successfully tested using multilevel analysis for the domains of mathematics and science. The results in relation to Hypotheses 1–4 (Models 2, 10, 4, and 12) suggest that the I/E model and the BFLPE are robust for both mathematics and science and demonstrate a successful extension of the two models from the typical domains of mathematics and verbal skills into the domain of science. Marsh and Shavelson's (1985) model predicts that academic self-concepts include two core academic domains (mathematics and English), while Shavelson et al.'s (1976) structure of self-concepts predicts that academic self-concepts include self-concepts in all school domains (e.g., mathematics, English, science, history, and the arts). The present findings suggest that mathematics and science can be distinctly different school subjects perceived by students through the psychological process of internal comparison in constructing their self-concepts in the two domains. The successful support for the I/E model and the BFLPE separately has also set the stage for combining them into one model using the same statistical procedure.

The major purpose of the present study was to justify and examine the combined model that integrates the I/E model and the BFLPE on the basis of the theoretical backgrounds and rationales for the two models with additional reference to theories and research on strong, reciprocal, and positive relations between achievements and self-concepts in matching domains at the individual level. The combined model was examined by the use of multilevel analysis for mathematics and science. The predictions of the basic I/E model (see Panels A and B in Figure 1) were supported when the effects of school achievements in the matching domains (the BFLPE) were included in a single model, a result supporting Hypotheses 5–6 (Models 7 and 15), even controlling for country achievements in the matching domains (Models 8 and 16). The results imply that the basic I/E and BFLPE models are solid models and robust theories, the predictions of which are not changed when relevant and significant variables are included in the models. On the other hand, the results also suggest that the basic I/E model may be an operation based on the process of internal comparison and that there may be a need to add the BFLPE (the process of external comparison) to the I/E model. Combining the I/E model and the BFLPE in a direct way addresses the issue that internal and external comparisons are distinctly different processes in constructing self-concept in a particular domain.

The first major claim of the combined model is that, although social comparison theory is one of the major theoretical backgrounds of the basic I/E model, the unique contribution of the basic I/E model to knowledge is actually internal comparison: Individuals' self-concept in a domain is influenced by comparison with their own performances in other domains. Relativity of their performances in different domains suggests that they measure their own performances in different domains on a single scale (i.e., internal frame of reference). There appears to be no need to assume that the strong effect of achievement on self-concept in matching domains is based on other frames of reference. The basic I/E model therefore can be viewed as providing a unique contribution to the creation of internal comparison as part of the psychological process in constructing academic self-concepts. Shavelson et al.'s (1976) conception of the multidimensional structure of self-concepts is well elaborated by the basic I/E model, with inspiration from later consistent findings that there are high correlations between mathematics and verbal achievements and low correlations between mathematics and verbal self-concepts. Research on rapid eye movement and achievement motivation has also insisted on and demonstrated a cause and effect relationship between individual achievement and self-concept, the conceptions of which all are tested at the individual level.

The second major claim of the combined model is that the operation of the BFLPE in past studies has precisely captured the essence of social comparison theory. Comparison with generalized others in a particular domain is in line with social comparison theory and multidimensional self-concepts, which forms the two theoretical backgrounds of the BFLPE. Schools or classes are a well-established, norm-based setting ready for external comparison. As such, school-average or class-average achievement can serve as a measure for evaluating the role of external comparison in constructing academic self-concept, while student achievement cannot. In other words, group achievement is a suitable measure for group effects, and student achievement is a suitable measure for individual effects. If the two measures are placed in the same model, the role of the two effects can be distinctly differentiated.

The combined model was tested on the basis of the these two major claims, and the results clearly distinguish the differential roles of internal and social comparisons in constructing self-concept in a particular domain. The process of internal comparison creates a phenomenon that student achievement leads to student self-concept in a positive direction for matching domains and leads to student self-concept in a negative direction for nonmatching domains. The process of external comparison creates a phenomenon that high-achieving schools/classes achievements lead to student self-concept in a negative direction and vice versa. The integration of the two processes in a combined model demonstrates that self-concept in a particular domain is simultaneously influenced by internal and external comparisons and that the meanings of internal and external comparisons can be clearly differentiated and tested with empirical data.

Implications for educational research and practice. An accurate theory of psychological processes in relation to the construction of academic self-concepts is of paramount importance for educational policy makers, teachers, and students. The clarification of the operations of the psychological processes of internal and external comparisons can facilitate the generation of appropriate teaching practices for students to develop proper self-concepts in different domains. What students need is not unrealistic confidence but a structure of sound and accurate multidimensional academic self-concepts obtained by internal and external comparisons in order to make sound career decisions and maximize achievements in society. The process of internal comparison (the basic I/E model) can help students enlarge the psychological distances between their self-concepts in different domains so that students can be aware of the relative strengths of their abilities, channel their development of unique abilities effectively, and best use their abilities in different domains creatively for different contexts. The process of external comparison (the BFLPE), on the other hand, can help students estimate their likelihood of unique contributions to society in particular domains. The combined model suggests simultaneous and distinctly different operations of the two processes.

The combined model also highlights a stronger effect of student achievement on student self-concept than the effect of group achievement on student self-concept in a particular domain in a single model. As such, accurate assessments of student achievements in different domains are still necessary for facilitation of the process of internal comparison. However, the idea of making assessment results a public issue (e.g., norm-reference tests or a streamed educational system) may need further consideration in terms of the BFLPE. There appears to be a need for teachers to increase students' knowledge of their relative strengths in diverse domains and highlight the importance of their unique patterns of abilities in different domains. There is also a need to remind students of their tendency to compare their achievements with their peers' in a setting and to give students opportunities to use a broader frame of reference in order to maximize their unique contributions to society (e.g., the use of teaching materials and activities based on a global perspective). Educators may also wish to know the relative weights that different students may place or should place on internal and external comparisons, by the understanding of which teachers can guide students into an adaptive construction of their multidimensional self-concept and maximize their unique contributions to society. These implications of the major findings may give some insights to teachers who intend to help students lower the negative influence by the BFLPE and develop appropriate self-concepts in addition to the teaching practices indicated by Bane, Haymaker, and Zinchuk (2005) and Lüdtke, Köller, and Marsh (2005).

Other Features, Limitations, and Implications for Future Research

Several minor features of the present study may need further consideration and have implications for future research.

Models in the domain of science. Few I/E model and BFLPE studies focus on the domain of science, and both of these models are supported for science in the present study. There is, however, a slight difference between the results for mathematics and science. In the combined model for science (see Model 15 in Table 3), the negative effect of student mathematics achievement on student science self-concept becomes nonsignificant. On the other hand, in the combined model for mathematics (see Model 7 in Table 2), the negative effect of student science achievement on student mathematics self-concept remains significant. This result may encourage researchers to take account of comparison between different domains of knowledge as an effective psychological process in constructing academic self-concept. For instance, the psychological distance between mathematics and science is likely to be different from that between mathematics and English.

Generalizability of the combined model and effect of country achievement. The posited combined model, especially that in predicting student mathematics self-concept, is supported for data from most countries. The present results are consistent with those of Marsh and Hau (2003) in relation to the BFLPE for average achievement from several domains, those of Marsh and Hau (2004) in relation to the I/E model for mathematics and verbal skills, and those of Chiu (2008) in relation to the I/E model for mathematics and science. On the other hand, there is significant variation in the degree of data fit to the combined model across countries, and the inclusion of country-level achievement in the combined model can reduce country-to-country variation. The present findings indicate that country achievement has significant effects on student self-concept in particular domains, which is in line with results of related studies (e.g., Leung, 2002, and Wilkins, 2004). The negative effect of country achievement may be a result of cultural backgrounds, pedagogical designs, or psychological process at the country level. The essence of the effect of distal contexts (e.g., countries) is worth investigation and elaboration in addition to the effect of the proximal context from classes and schools (i.e., the BFLPE).

Statistical methods. Multilevel analysis is not a typical procedure for examining the I/E model, but it is for examining the BFLPE. Multilevel analysis is a statistical method that takes into account variances within both individual and group levels, which appropriately meets the structure of most sampling procedures in large-scale tests (e.g., TIMSS). Future research, however, can still aim to develop advanced techniques for modeling the combined model. Four suggestions are as follows:

1. Statistical techniques that integrate latent-variable, multilevel, and path modeling: Recent research on multilevel effects has suggested that the group-level aggregate is an estimate of a true population mean that contains measurement errors at the individual level and sampling and measurement errors at the group level. The size of the bias due to this problem varies with the number of student in each group, the number of groups, the sampling ratio, and the nature of the aggregated variable (Lüdtke et al., 2008). Most measurement errors can be considered, controlling for using latent-variable modeling. Most sampling errors may be considered using multilevel modeling with sampling weights. Path analysis can model data in a cause-and-effect manner, like the I/E model.

2. Models that include cross-level interactions: Another way to improve the combined model is to include the effects of cross-level interactions. This approach was initiated by Marsh et al. (2009), who investigated the cross-level relationships between individual and group variables in their study of the BFLPE. The effects of interactions between student and group achievements on academic self-concept may be an interesting topic for future research, as suggested by Marsh and Hau (2003) and Wilkins (2004).

3. Sampling that can solve the problem of multicollinearity: A much better combined model may need to allow for the inclusion of both school mathematics and science achievements in a model to investigate their BFLPEs on mathematics and science self-concepts, respectively. However, this inclusion may create the

problem of multicollinearity, as the correlations between group-level achievements in different domains are generally large. In other words, data that have low correlations between group-level achievements are needed. A way to solve this problem may be to use data from more distinctly different domains (e.g., arts and engineering) or from participants who show significantly differential achievements between two different domains (e.g., gifted students good at mathematics but bad at some other domains).

4. Qualitative research for elaborating the combined model: In-depth interviews with significant participants sensitive to internal and external/social comparisons in particular contexts (e.g., students with special gifts who have experiences in both gifted and normal classes) may provide a full picture of the complex issues underlying the combined model. Qualitative research methodologies are likely to create a new combined model, beyond the basic configuration of the I/E model and the BFLPE. Later quantitative research can be used to validate the new combined model.

The present study used a single wave of data to examine the likely "causal" relationship (e.g., regressing self-concepts on achievements), but only results based on analyses of a number of waves of data can be better used to make a claim of causal relationships (cf. Guay et al., 2003; Marsh et al., 2000; Marsh, Köller, & Baumert, 2001; Marsh et al., 2002). While only results obtained by experimental methods can be utilized to make a claim of cause and effect, the use of experimental methods for studying affective issues may be relatively unethical in the case of young students as research participants. Therefore, survey data still appear to be one of the more appropriate sources for future research to address similar issues.

References

Abu-Hilal, M. M., & Bahri, T. M. (2000). Self-concept: The generalizability of research on the SDQ, Marsh/Shavelson model and I/E frame of reference model to United Arab Emirates students. *Social Behavior and Personality, 28,* 309–322. doi:10.2224/sbp.2000.28.4.309

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Bane, C. M. H., Haymaker, C. M. B., & Zinchuk, J. (2005). Social support as a moderator of the big-fish-in-a-little-pond effect in online self-help support groups. *Journal of Applied Biobehavioral Research, 10,* 239–261. doi:10.1111/j.1751-9861.2005.tb00015.x

Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. *Journal of Educational Psychology, 90,* 102–110. doi:10.1037/0022-0663.90.1.102

Burleson, K., Leach, C. W., & Harrington, D. M. (2005). Upward social comparison and self-concept: Inspiration and inferiority among art students in an advanced program. *British Journal of Social Psychology, 44,* 109–123. doi:10.1348/014466604X23509

Burton, L. (1994). *Children learning mathematics: Patterns and relationships.* Hemel Hempstead, England: Simon & Schuster Education.

Cheng, R. W. Y., & Lam, S. F. (2007). Self-construal and social comparison effects. *British Journal of Educational Psychology, 77,* 197–211. doi:10.1348/000709905X72795

Chiu, M. S. (2008). Achievements and self-concepts in a comparison of mathematics and science: Exploring the internal/external frame of reference model across 28 countries. *Educational Research and Evaluation, 14,* 235–254. doi:10.1080/13803610802048858

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

Du Toit, M., & Du Toit, S. H. C. (2001). *Interactive LISREL: User's guide.* Lincolnwood, IL: Scientific Software International.

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

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology, 33,* 9–33. doi:10.1016/j.cedpsych.2006.12.002

Gonzalez, E. J., Galia, J., Arora, A., Erberber, E., & Diaconu, D. (2004). Reporting student achievement in mathematics and science. In M. O. Martin, I. V. S. Mullis, & S. J. Chrostowski (Eds.), *TIMSS 2003 technical report* (pp. 275–307). Chestnut Hill, MA: TIMSS & PIRLS International Study Centre.

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

Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Helmke, A., & Van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology, 87,* 624–637. doi:10.1037/0022-0663.87.4.624

Hoge, D. R., Smit, E. K., & Crist, J. T. (1995). Reciprocal effects of self-concept and academic achievement in sixth and seventh grade. *Journal of Youth and Adolescence, 24,* 295–314. doi:10.1007/BF01537598

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

International Association for the Evaluation of Educational Achievement. (2005). *TIMSS 2003 user guide for the international database.* Chestnut Hill, MA: TIMSS & PIRLS International Study Centre.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8: User's reference guide.* Lincolnwood, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (2005). *LISREL* (Version 8.72) [Computer software]. Lincolnwood, IL: Scientific Software International.

Kemmelmeier, M., & Oyserman, D. (2001). The ups and downs of thinking about a successful other: Self-construals and the consequences of social comparisons. *European Journal of Social Psychology, 31,* 311–320. doi:10.1002/ejsp.47

Kurtz-Costes, B. E., & Schneider, W. (1994). Self-concept, attributional beliefs, and school achievement: A longitudinal analysis. *Contemporary Educational Psychology, 19,* 199–216. doi:10.1006/ceps.1994.1017

Leung, F. K. S. (2002). Behind the high achievement of East Asian students. *Educational Research and Evaluation, 8,* 87–108. doi:10.1076/edre.8.1.87.6920

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,* 567–586. doi:10.1348/000709905X42239

Lüdtke, O., Köller, O., & Marsh, H. W. (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. (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

Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education, 110,* 256–280. doi:10.1086/383074

Manger, T., & Eikeland, O. J. (1997). The effect of social comparison on mathematics self-concept. *Scandinavian Journal of Psychology, 38,* 237–241. doi:10.1111/1467-9450.00032

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

Marsh, H. W. (1986). Verbal and mathematics self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23,* 129–149.

Marsh, H. W. (1987a). 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. (1987b). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement, 24,* 17–39. doi:10.1111/j.1745-3984.1987.tb00259.x

Marsh, H. W. (1989). Sex differences in the development of verbal and mathematics constructs: The High School and Beyond study. *American Educational Research Journal, 26,* 191–225.

Marsh, H. W. (1990a). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology, 82,* 107–116. doi:10.1037/0022-0663.82.1.107

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

Marsh, H. W. (1990c). The structure of academic self-concept: The Marsh/Shavelson Model. *Journal of Educational Psychology, 82,* 623–636. doi:10.1037/0022-0663.82.4.623

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

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal, 30,* 841–860.

Marsh, H. W. (1994). Using the national 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., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80,* 366–380. doi:10.1037/0022-0663.80.3.366

Marsh, H. W., Chanal, J. P., & Sarrazin, P. G. (2006). Self-belief does make a difference: A reciprocal effects model of the causal ordering of physical self-concept and gymnastics performance. *Journal of Sports Sciences, 24,* 101–111. doi:10.1080/02640410500130920

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. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science, 1,* 133–163. doi:10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Craven, R., & Debus, R. (1999). Separation of competency and affect components of multiple dimensions of academic self-concept: A developmental perspective. *Merrill-Palmer Quarterly, 45,* 567–601.

Marsh, H. W., Craven, R. G., & McInerney, D. M. (Eds.). (2003). *International advances in self research.* Greenwich, CT: Information Age.

Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 38–90). New York, NY: Wiley.

Marsh, H. W., & Hau, K. T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist, 58,* 364–376. doi:10.1037/0003-066X.58.5.364

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96,* 56–67. doi:10.1037/0022-0663.96.1.56

Marsh, H. W., Hau, K. T., & Kong, C. K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal, 39,* 727–763. doi:10.3102/00028312039003727

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

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

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

Marsh, H. W., Kong, C. K., & Hau, K. T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology, 93,* 543–553. doi:10.1037/0022-0663.93.3.543

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

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

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

Marsh, H. W., & Roche, L. A. (1996). Structure of artistic self-concepts for performing arts and non-performing arts students in a performing arts high school: "Setting the stage" with multigroup confirmatory factor analysis. *Journal of Educational Psychology, 88,* 461–477. doi:10.1037/0022-0663.88.3.461

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

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

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

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores:

Reciprocal effects models of causal ordering. *Child Development, 76,* 397–416. doi:10.1111/j.1467-8624.2005.00853.x

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

Marsh, H. W., & Yeung, A. S. (1997a). 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

Marsh, H. W., & Yeung, A. S. (1997b). Coursework selection: The effects of academic self-concept and achievement. *American Educational Research Journal, 34,* 691–720.

Marsh, H. W., & Yeung, A. S. (2001). An extension of the internal/external frame of reference model: A response to Bong (1998). *Multivariate Behavioral Research, 36,* 389–420. doi:10.1207/S15327906389-420

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international science report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Centre.

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

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

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

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education, 21,* 467–487. doi:10.1007/BF03173515

Mui, F. L. L., Yeung, A. S., Low, R., & Jin, P. (2000). Academic self-concept of talented students: Factor structure and applicability of the internal/external frame of reference model. *Journal for the Education of the Gifted, 23,* 343–367.

Muijs, R. D. (1997). Predictors of academic achievement and academic self-concept: A longitudinal perspective. *British Journal of Educational Psychology, 67,* 263–277. doi:10.1111/j.2044-8279.1997.tb01243.x

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research, 151,* 53–79. doi:10.1016/S0377-2217(02)00578-7

Plucker, J. A., & Stocking, V. B. (2001). Looking outside and inside: Self-concept development of gifted adolescents. *Exceptional Children, 67,* 535–548.

Pottebaum, S. M., Keith, T. Z., & Ehly, S. W. (1986). Is there a causal relation between self-concept and academic achievement? *Journal of Educational Research, 79,* 140–144.

Preckel, F., Zeidner, M., Goetz, T., & Schleyer, E. J. (2008). Female "big fish" swimming against the tide: The "big-fish-little-pond effect" and gender-ratio in special gifted classes. *Contemporary Educational Psychology, 33,* 78–96. doi:10.1016/j.cedpsych.2006.08.001

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

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., & Du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Raudenbush, S., Bryk, A., & Congdon, R. (2005). *HLM for Windows* (Version 6.02a) [Computer software]. Lincolnwood, IL: Scientific Software International.

Rogers, C. M., Smith, M. D., & Coleman, J. M. (1978). Social comparison in the classroom: The relationship between academic achievement and self-concept. *Journal of Educational Psychology, 70,* 50–57. doi:10.1037/0022-0663.70.1.50

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

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London, England: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. doi:10.1037/1082-989X.7.2.147

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

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology, 58,* 292–307. doi:10.1037/0022-3514.58.2.292

Skaalvik, E. M., & Rankin, R. J. (1992). Mathematics and verbal achievement and self-concepts: Testing the internal/external frame of reference model. *Journal of Early Adolescence, 12,* 267–279. doi:10.1177/0272431692012003003

Skaalvik, E. M., & Rankin, R. J. (1995). A test of the internal/external frame of reference model at different levels of mathematics and verbal self-perception. *American Educational Research Journal, 32,* 161–184.

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist, 37,* 233–244. doi:10.1207/S15326985EP3704_3

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

Stapleton, L. M. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling, 9,* 475–502. doi:10.1207/S15328007SEM0904_2

Suls, J. M. (1977). Social comparison theory and research: An overview from 1954. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 1–19). New York, NY: Halsted Press.

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

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

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

Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *Journal of Experimental Education, 72,* 331–346. doi:10.3200/JEXE.72.4.331-346

Williams, J. E., & Montgomery, D. (1995). Using frame of reference theory to understand the self-concept of academically able students. *Journal for the Education of the Gifted, 18,* 400–409.

Wong, M. S. W., & Watkins, D. (2001). Self-esteem and ability grouping: A Hong Kong investigation of the big fish little pond effect. *Educational Psychology, 21,* 79–87. doi:10.1080/01443410123082

Yeung, A. S., & Lee, F. L. (1999). Self-concept of high school students in

China: Confirmatory factor analysis of longitudinal data. *Educational and Psychological Measurement, 59*, 431–450. doi:10.1177/00131649921969965

Zeidner, M., & Schleyer, E. 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

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511527692.009

Received October 26, 2008
Revision received August 16, 2011
Accepted September 6, 2011 ■