

The Murky Distinction Between Self-Concept and Self-Efficacy: Beware of Lurking Jingle-Jangle Fallacies

Herbert W. Marsh

Australian Catholic University and Oxford University

Reinhard Pekrun

University of Munich and Australian Catholic University

Philip D. Parker

Australian Catholic University

Kou Murayama

University of Reading and Kochi University of Technology

Jiesi Guo and Theresa Dicke

Australian Catholic University

A. Katrin Arens

German Institute for International Educational Research

This study extends the classic constructive dialogue/debate between self-concept and self-efficacy researchers (Marsh, Roche, Pajares, & Miller, 1997) regarding the distinctions between these 2 constructs. The study is a substantive-methodological synergy, bringing together new substantive, theoretical, and statistical models and developing new tests of the classic jingle-jangle fallacy. We demonstrate that in a representative sample of 3,350 students from math classes in 43 German schools, generalized math self-efficacy and math outcome expectancies were indistinguishable from math self-concept, but were distinct from test-related and functional measures of self-efficacy. This is consistent with the jingle-jangle fallacies that are proposed. On the basis of pretest variables, we demonstrate negative frame-of-reference effects in social (big-fish-little-pond effect) and dimensional (internal/external frame-of-reference effect) comparisons for three self-concept-like constructs in each of the first 4 years of secondary school. In contrast, none of the frame-of-reference effects were significantly negative for either of the two self-efficacy-like constructs in any of the 4 years of testing. After controlling for pretest variables, each of the 3 self-concept-like constructs (math self-concept, outcome expectancy, and generalized math self-efficacy) in each of the 4 years of secondary school was more strongly related to posttest outcomes (school grades, test scores, future aspirations) than were the corresponding 2 self-efficacy-like factors. Extending discussion by Marsh et al. (1997), we clarify distinctions between self-efficacy and self-concept; the role of evaluation, worthiness, and outcome expectancy in self-efficacy measures; and complications in generalized and global measures of self-efficacy.

Educational Impact and Implications Statement

Positive self-beliefs are a central construct in educational psychology, and self-concept and self-efficacy are the most widely used and theoretically important representations of positive self-beliefs. In Educational Psychology, much effort has been expended in trying to distinguish between self-concept and self-efficacy. Nevertheless, in practice and theory the distinction remains murky. We critique previous conceptual attempts to distinguish the two constructs—arguing against some distinctions that have been offered in the past and offering some new theoretical distinctions and new empirical approaches to testing support for these distinctions.

Keywords: self-concept, self-efficacy, social comparison, dimensional comparison, jingle-jangle fallacy

Supplemental materials: <http://dx.doi.org/10.1037/edu0000281.supp>

This article was published Online First July 2, 2018.

Herbert W. Marsh, Institute of Positive Psychology and Education, Australian Catholic University, and Department of Education, Oxford University; Reinhard Pekrun, Department of Psychology, University of Munich, and Institute for Positive Psychology and Education, Australian Catholic University; Philip D. Parker, Institute of Positive Psychology and Education, Australian Catholic University; Kou Murayama, School of Psychology and Clinical Language Sciences, University of Reading, and Research Institute, Kochi University of Technology; Jiesi Guo and Theresa Dicke, Institute of Positive Psychology and Education, Australian Catholic University; A. Katrin Arens, Education and Human Development, German Institute for International Educational Research.

This research is in part supported by a grant from the Australian Research Council (DP130102713) awarded to Herbert W. Marsh, four grants from the German Research Foundation (DFG) awarded to Reinhard Pekrun (PE 320/11-1, PE 320/11-2, PE 320/11-3, PE 320/11-4), and a grant from the Japan Society of Promoting Science (KAKENHI grant 16H06406) awarded to Kou Murayama.

Correspondence concerning this article should be addressed to Herbert W. Marsh, Institute of Positive Psychology and Education, Australian Catholic University, 33 Berry Street, North Sydney, NSW 2060. E-mail: Herb.Marsh@acu.edu.au

Positive self-beliefs, dating back at least to William James (1890/1963; Marsh, 2007) but arguably to Socrates and Plato (see Hattie, 1992), are among the oldest and most widely studied psychological constructs. Self-beliefs are central in theoretical models of motivation, as well as in psychological theories more generally. Thus, Elliot and Dweck (2005; also see Marsh, Martin, Yeung, & Craven, 2017) concluded that competency self-perceptions were all-pervasive and powerful:

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

Marsh and colleagues (Marsh & Craven, 2006; Marsh, Martin, et al., 2017) have argued that self-beliefs are central to the positive psychology movement. In recognition of their importance, the enhancement of positive self-beliefs is identified as a major focus of concern in diverse settings, including education, child development, mental and physical health, and the social sciences more generally. However, this broad popularity and multidisciplinary appeal also comes at a cost in terms of construct definition, measurement, validation, and rigor. With so many researchers from so many disciplines measuring self-belief constructs, inevitably a plethora of similarly labeled constructs have arisen that denote different phenomena, as well as differently labeled constructs that denote similar phenomena.

Self-concept and self-efficacy are the most widely used and theoretically important representations of positive self-beliefs. In this article, focused on the murky distinction between these two constructs, we reintroduce Kelley's (1927; Marsh, 1994) classic jingle-jangle fallacy and provide a construct-validation framework to test for this fallacy that has wide applicability to psychological measurement, theory, and practice. On the basis of the nature and construction of items used to infer the constructs, we posit an a priori classification of diverse self-belief constructs as either self-concept-like or self-efficacy-like constructs. We empirically test this theoretical classification on the basis of relations among factors using the logic of multitrait-multimethod analysis, classic frame-of-reference effects (social and dimensional comparison effects) that influence self-concept formation but are posited to be attenuated for self-efficacy responses, and long-term predictions of critical outcomes (grades, test scores, aspirations) from four waves of self-belief measures—after controlling for preexisting differences.

Jingle-Jangle Fallacies and Construct Validation

Researchers have conceptualized positive self-beliefs from a variety of theoretical perspectives (self-concept, self-esteem, self-efficacy, expectations of success, agency, locus of control, outcome expectations, confidence, competency, growth mind-set, etc.; see Skinner, 1996, for similar problems with constructs of control). Particularly in studies of self-beliefs and motivation more generally, researchers tend to focus on their preferred measures, sometimes paying relatively little attention to testing how (or whether) they differ from other, apparently related constructs (see related discussion by Marsh, Craven, Hinkley, & Debus, 2003; Murphy & Alexander, 2000; Parker, Marsh, Ciarrochi, Marshall,

& Abduljabbar, 2014; Seifert, 2004). This leads to jingle-jangle fallacies (Block, 1995; Marsh, 1994; Marsh et al., 2003), a phrase first coined by Kelley (1927); two scales with similar names might measure different constructs (jingle fallacy) while two scales with apparently dissimilar labels might measure similar constructs (jangle fallacy).

Marsh (1994) demonstrated jingle-jangle fallacies in a factor analysis of two different motivation instruments. He found that mastery and performance scales from each instrument reflected common underlying factors. However, the competition scales from the instruments reflected different constructs (a performance orientation and a task orientation), even though they had the same label. To test (and avoid) jingle-jangle fallacies, researchers need to conduct construct validity studies to test interpretations of the measures. Indeed, at the level of items, a finding that items from a given scale load on a single factor when only that one scale is considered does not test whether the items will load on different factors when different constructs are included in a single factor analysis. At the level of scales, the label assigned to a factor is not a sufficient basis for establishing how that scale relates to other, apparently similar or dissimilar constructs. Heyman and Dweck (1992) similarly noted that researchers “need to take care that they are not measuring the same construct disguised in different scale names” (p. 243). Pajares (2009) noted problems with conceptually similar measures that are differentially operationalized to suit different research agendas, leaving researchers the task of sorting through the different measures; his particular concern was researchers inappropriately labeling competence perceptions as self-efficacy perceptions. Similarly, Bong (1996; Bong & Skaalvik, 2003) suggested that to avoid a “conceptual mess,” researchers should apply confirmatory factor analysis (CFA) and structural equation models (SEMs) to evaluate the structural, predictive, convergent, and discriminant validity of different motivation measures. Thus, more emphasis on convergent and discriminant validity across multiple constructs and the application of statistical tools such as CFA, SEM, and multitrait-multimethod (MTMM) analysis is needed.

The Murky Distinction Between Self-Concept, Self-Efficacy, and Outcome Expectancy

In this introduction, we develop theoretical distinctions between self-concept, self-efficacy, and outcome expectancy and discuss the relevance of these distinctions in applied research. We begin with a brief overview of research on the formation of academic self-concept, with a particular emphasis on frame-of-reference effects, which are a major focus of our study. We then juxtapose this self-concept research with self-efficacy research in a brief review of similarities and distinctions between the two. Next, we discuss how outcome expectancy is related to self-concept and self-efficacy. The review of self-concept theory and research focuses on frame-of-reference effects, which have been the basis of much recent self-concept research. Then, in our review of self-efficacy theory and research, we argue that appropriately designed self-efficacy items should largely eliminate such frame-of-reference effects, and we explore the implications of this proposal. Finally, we draw upon the literature to propose a priori hypotheses to clarify the murky distinction between self-concept, self-efficacy, and outcome expectancy.

Self-Concept Theory and Research

In the last quarter century, self-concept research has seen a resurgence in the quality and sophistication of theoretical models, research design, quantitative methodology, and measurement instruments. This was stimulated in part by Shavelson, Hubner, and Stanton's (1976) seminal review article, which evaluated existing self-concept research and developed a new multidimensional, hierarchical model of self-concept that was the basis of new multidimensional self-concept instruments for the next generation and beyond (see review by Marsh & Hattie, 1996; Marsh & Shavelson, 1985). Integrating key features from the 17 different conceptual definitions of self-concept identified, Shavelson et al. broadly defined self-concept as a person's self-perceptions formed through experience with and interpretations of his or her environment. These included feelings of self-confidence, self-worth, self-acceptance, competence, and ability. They noted that self-concept is influenced especially by the evaluations of significant others, by reinforcements, and by attributions for one's behavior. These self-perceptions influence the way one acts, and these acts in turn influence one's self-perceptions.

From as early as William James (1890/1983), psychologists have emphasized that self-concepts are based on objective accomplishments evaluated in relation to frames of reference or standards of comparison. Self-evaluations of competence in a particular domain can be made against many different frames of reference or standards of comparison (Marsh & Seaton, 2015; Skaalvik & Skaalvik, 2002): an absolute ideal (e.g., the 5-min mile), social comparisons (e.g., results of classmates on a test), temporal comparisons (e.g., improvement over time, a personal best), or dimensional comparisons (e.g., one's accomplishments in one domain

relative to one's own accomplishments in other domains). Theoretical models of how such frame-of-reference effects influence self-concept have been a major focus of recent research, particularly in relation to academic self-concept. However, there is much theoretical and empirical confusion about the role of frame-of-reference effects in relation to self-efficacy responses. This lies at the heart of the murky distinctions between the two constructs and is the major focus of the present investigation.

Internal/external frame-of-reference (I/E) model: Dimensional comparison effects. Academic self-concepts (ASCs) in specific school subjects are much more differentiated than are the corresponding measures of achievement. Thus, verbal and math achievements tend to be substantially correlated, but verbal and math self-concepts tend to be nearly uncorrelated (Marsh, 1986, 2007; Marsh, Kuyper, Seaton, et al., 2014; Marsh, Xu, & Martin, 2012; Möller & Marsh, 2013). Providing a theoretical rationale for these results, the I/E model posits that ASCs in a particular school subject are formed relative to two frames of reference: an external (social comparison) reference based on comparisons of one's performances with those of other students in the same school subject, and an internal (dimensional comparison) reference based on one's own performance in that school subject with one's own performances in other school subjects.

According to the I/E model (Figure 1A), achievement is substantially related to ASC in the same (matching) domain. However, the key theoretical prediction is that the cross-paths leading from achievement in one domain to ASCs in a different (non-matching) domain (e.g., verbal achievement to math self-concept) are negative. The rationale for this prediction is that students will use verbal achievement, for example, as a basis for comparison in

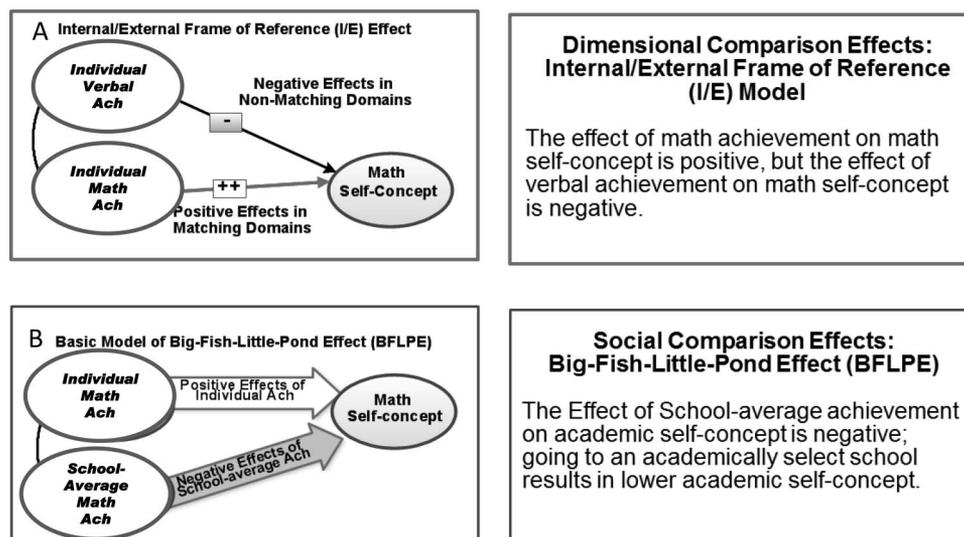


Figure 1. Schematic diagram of two frame-of-reference effects evaluated in the present investigation. In both models, the effect of individual student math achievement on math self-concept is positive. In the internal/external frame of reference (I/E) model (Figure 1A) there is a negative (dimensional comparison) effect of verbal achievement on math self-concept. In the big-fish-little-pond effect model (Figure 1B) there is a negative frame-of-reference (social comparison) effect of school-average-achievement on math self-concept. These two effects are posited to work simultaneously; math self-concept is formed in relation to dimensional comparisons (accomplishments in one domain relative to one's accomplishments in other domains) and social comparisons (one's accomplishments relative to those in one's peer group).

the formation of their math self-concept. Thus, good verbal achievement will lead to good verbal self-concept, but actually detract from a high math self-concept. Using Programme for International Student Assessment (PISA) data, Marsh & Hau (2004) showed that support for the I/E model predictions generalized over 26 countries. Subsequently, the Möller, Pohlmann, Köller, and Marsh (2009) meta-analysis similarly found that although math and verbal achievements were highly correlated ($r = .67$), math and verbal self-concepts were nearly uncorrelated ($r = .10$). The path analysis based on this meta-analytic data showed that the paths leading from math achievement to math self-concept were substantially positive ($\beta = .61$). However, paths leading from verbal achievement to math self-concept were negative ($\beta = -.27$), consistent with the I/E model.

The big-fish-little-pond effect (BFLPE): Social comparison effects. The BFLPE model (Figure 1B) posits that students compare their own academic abilities with those of their classmates and use this social comparison to form their ASCs (Marsh, Abduljabbar, et al., 2015; Marsh, Kuyper, Morin, et al., 2014; Marsh & Parker, 1984; Marsh, Seaton, et al., 2008; Nagengast & Marsh, 2012; Parker, Marsh, et al., 2018; Tymms, 2001; Zell & Alicke, 2009). According to the BFLPE, students in schools that have a high school-average achievement will have lower ASCs than will equally able students in schools where the school-average ability is not high; as such, school-average achievement has a negative effect on ASC.

Much support has been found for the BEFLPE (see review by Marsh, Seaton, et al., 2008). In results based on three successive PISA data collections (Marsh & Hau, 2003: 103,558 students from 26 countries; Seaton, Marsh, & Craven, 2010: 265,180 students from 41 countries; Nagengast & Marsh, 2012: 397,500 students from 57 countries), the effects of school-average achievement on ASC were negative in 122 of 123 country samples, and significantly so in 114 samples. In addition, the BFLPE tends to increase over time when students attend the same high school (Marsh, Köller, & Baumert, 2001). Furthermore, Marsh et al. (2007) have shown that the BFLPE formed in high school is maintained 2 and 4 years after high school. It is important that, apart from ASC, the BFLPE has been shown to have a negative effect on many other desirable educational outcomes, including educational aspirations, general self-concept, school grades, standardized test scores, advanced coursework selection, subsequent university attendance, and occupational aspirations (Marsh, 1991). Furthermore, these negative effects of school-average achievement on a range of other constructs were at least partially mediated by ASC.

Self-Efficacy: What It Is and how It Differs From Self-Concept

According to Bandura (1994, p. 71), “Perceived self-efficacy is defined as people’s beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self-efficacy beliefs determine how people feel, think, motivate themselves and behave.” A critical feature of self-efficacy theory is that it distinguishes between motivation to perform a target behavior and self-perceptions of capability to perform the behavior. As emphasized by Bong and Skaalvik (2003) and others (e.g., Marsh, 2007; Parker, Marsh, Lüdtke, & Trautwein, 2013; Schunk & Pajares, 2005; Zimmerman,

2000), academic self-efficacy and academic self-concept have much in common: an emphasis on perceived competence, a multidimensional and hierarchical structure, content specificity, and the prediction of future performance, emotion, and motivation.

Historically, self-concept was argued to be a global construct, whereas self-efficacy was a very domain- and task-specific construct (Bandura, 1986). However, in current theoretical models of self-concept, self-concept facets are as domain-specific as are typical self-efficacy measures, while some self-efficacy researchers focus on generalized measures of self-efficacy (e.g., General Self-Efficacy Scale; Schwarzer & Jerusalem, 1995; Motivated Strategies for Learning Questionnaire; Pintrich & De Groot, 1990). Nevertheless, self-efficacy researchers have neither developed nor tested multidimensional, hierarchical models of self-efficacy that integrate global and increasingly specific components of self-efficacy such as those underlying self-concept theory. Indeed, Maddux (2009) suggests that global and generalized measures of self-efficacy are less useful than more specific measures, and posits their continued use as an unresolved issue for further research. Hence, in relation to globality, the distinction between self-efficacy and self-concept does not appear to be very useful. For the present purposes we focus on three key characteristics that distinguish self-efficacy from self-concept.

Prospective versus retrospective. The first distinguishing feature is that self-efficacy responses are constructed to be prospective: They address what one is able to accomplish in the future in relation to a specific task in a particular context. Indeed, this is why Bandura (1986) emphasized self-efficacy expectations. Hence, Bandura (1986, 1989, 1997) and others (e.g., Schunk & Pajares, 2005) suggest that self-efficacy refers to beliefs about “what I can do”: cognitive, goal-referenced, relatively context-specific, future-oriented judgments in relation to a narrowly defined task (Bong & Skaalvik, 2003; Schunk & Pajares, 2005). In contrast, although self-concept is predictive of future behavior and outcomes, it is largely based on past accomplishments and circumstances. Hence, a critical, unresolved issue is how well domain-specific measures of self-concept and self-efficacy predict future performance in longitudinal studies, after controlling for preexisting differences.

In self-efficacy research, a frequently used paradigm is to compare the ability of self-efficacy measures to predict test scores (Schunk & Pajares, 2005) that are typically administered in the same testing session. The self-efficacy items in this paradigm are similar to or, perhaps, the same as the actual test items that are subsequently presented (hereafter we refer to this as *test-related self-efficacy*). Consistently with the specificity matching principle (Pajares & Miller, 1995; also see Brunswick, 1952), it is not surprising that self-efficacy measures predict test performance based on similar or same test items, better than more generic self-concept measures. However, a more relevant test would be to first control for preexisting differences, and then test how well self-efficacy and self-concept predict a broader range of future performance and behavior that is temporally more removed and less directly tied to the specific self-efficacy items used (Marsh et al., 1997; Parker, Marsh, et al., 2013). Thus, the Valentine, DuBois, and Cooper (2004; also see Huang, 2011) meta-analysis of longitudinal studies of reciprocal effects models showed that self-belief constructs predicted future academic achievement even after controlling for prior achievement. However, they found that

there were no differences between domain-specific academic self-concept and self-efficacy measures, although both predicted subsequent achievement better than did more generalized measures such as self-esteem. However, because only one of the longitudinal studies in their meta-analysis included measures both of academic self-concept and self-efficacy, it did not offer a strong test of this distinction.

Subsequently, Huang (2012) conducted a meta-analysis and systematic review of the discriminant and incremental validity of self-concept and academic self-efficacy. Based on 74 mostly cross-sectional studies, the mean correlation between self-concept and self-efficacy was .43, and higher when the domain specificity of the two constructs matched. Their meta-analysis suggested that self-efficacy had higher incremental validity than self-concept. However, their secondary analyses of three waves of PISA data showed more nuanced results: Self-concept had higher incremental validity for prediction of school grades, but self-efficacy had more incremental validity in relation to PISA test scores. Huang (2012, p. 799) cautioned that “researchers need to realize that the wording and domain specificity of self-measures, as well as domain matching of self-measures and academic achievement, affect predictive power.” However, Huang operationalized incremental validity in relation to how much one of the constructs—self-concept or self-efficacy—was able to predict, after controlling for the effects of the other. An alternative perspective on incremental validity, the focus of the present investigation, is how much either construct is able to add to the prediction of a range of posttest measures after controlling for pretest differences in background demographic variables—including prior achievement. Not surprisingly, stronger tests of incremental validity are possible when based on longitudinal panel designs with multiple waves of data that provide stronger controls for preexisting differences and change over time.

Descriptive versus evaluative. A second distinguishing feature is that appropriately constructed self-efficacy responses are designed to be purely descriptive, whereas self-concept responses are both descriptive and evaluative. Thus, as emphasized by Bong and Skaalvik (2003) and others, paradigmatic, appropriately constructed self-efficacy items “solicit goal referenced evaluations and do not directly ask students to compare their abilities to those of others” (p. 9) and “provide respondents with a specific description of the required referent against which to judge their competence” (p. 9), whereas “assessing one’s capability in academic self-concept relies heavily on social comparison information” (p. 9). Similarly, Bandura (1986) argued that self-esteem and self-concept—but not self-efficacy—are partly determined by “how well one’s behavior matches personal standards of worthiness” (p. 410). Likewise, Pajares (2009, p. 546) distinguished self-efficacy from “Assessments of other expectancy beliefs include asking students to report how well they expect to do in an academic subject (i.e., performance expectancies, Meece, Wigfield, & Eccles, 1990), whether they understand what they read (i.e., perceptions of competence, Harter, 1986), and whether they are good in an academic subject (i.e., academic domain-specific self-concept, Marsh & Shavelson, 1985; also ability perceptions, Meece et al., 1990).” Hereafter, we refer to measures constructed according to these paradigmatic principles as relatively “pure” self-efficacy measures. Thus, for example, in a typical operationalization of self-efficacy, students are shown example math test items and asked for their confidence of correctly answering such items; their

responses are based on an absolute criterion that does not require them to compare their own performances with those of other students (also see Bong & Skaalvik, 2003).

In discussion of this distinction between self-concept and self-efficacy, Marsh (2007) argued that much of the power of self-beliefs to motivate and predict future behavior depends on the evaluation one makes of a pure performance expectation. Whereas the self-efficacy belief that I can run 100 m in 13 s in the next school track might be descriptive in nature, the self-evaluation of this outcome—whether this represents a great result or a terrible one—has important implications. Relatedly, Bong and Clark (1999) acknowledge that “self-concept is judged to be more inclusive . . . because it embraces a broader range of descriptive and evaluative inferences with ensuing affective reactions” (p. 142). Hence, even though carefully defined pure self-efficacy measures might be more future-oriented, and self-concept based more on past performance, after controlling for preexisting differences, self-concept should be able to predict a broader range of future performance and choice behaviors better than self-efficacy measures, particularly outcomes not directly tied to the specific content of the self-efficacy items.

Related to this distinction, Bandura (1977, 1986, 1994) has consistently argued that self-efficacy is distinct from and causally precedes outcome expectations of success or failure. However, Williams and Rhodes (2016; also see Eastman & Marzillier, 1984) offered a series of experimental studies showing that outcome expectations do influence self-efficacy. Furthermore, they showed that typical self-efficacy items conflate self-efficacy and motivation (will-do motivation vs. can-do capability), positively biasing measures of self-efficacy in relation to predicting outcomes. Thus, more carefully constructed items that were more purely self-efficacy (“I can do”) items substantially reduced the predictive validity of self-efficacy responses. However, Williams and Rhodes suggested that the problem might be with the measures of self-efficacy more than the construct itself. Nevertheless, that research resulted in the distinction between self-concept and self-efficacy becoming even murkier and emphasized again the critical problems associated with the appropriate construction of self-efficacy items.

Frame-of-reference effects. A third distinguishing feature is frame-of-reference effects, which have been so important in recent studies of self-concept formation. Theoretically, these effects should be largely eliminated in appropriately constructed (“pure”) self-efficacy responses. Thus, Marsh (2007) proposed that both the BFLPE and the I/E model should be substantially attenuated for responses to pure self-efficacy items, relative to frame-of-reference effects associated with self-concept responses. This distinction was highlighted in early research on the I/E model by Skaalvik and Rankin (1990), who purported to demonstrate that the model did not work for Norwegian students. However, Marsh, Walker, and Debus (1991) subsequently noted that the Skaalvik and Rankin study used (what here we refer to as) test-related self-efficacy measures that were consistent with Bandura’s original design guidelines. In particular, students were shown test items like those on the test, rather than the typical math and verbal self-concept scales used to develop the I/E model, and asked how likely they were to be able to answer test items of this type. Marsh et al. (1991) subsequently tested this distinction on the basis of test-related self-efficacy, self-concept, and test scores in the verbal and

math domains. Consistently with predictions, there was strong support for the I/E model based on self-concept measures, but none for test-related self-efficacy measures. Marsh (2007; Marsh, Trautwein, Lüdtke, & Köller, 2008) proposed that a similar logic should apply to the BFLPE. Thus, being in an academically selective school with other academically gifted classmates should not have much effect on academic pure self-efficacy but should have a negative effect on academic self-concept. Although the research is sparse and there is apparently no strong empirical support for this theory-based hypothesis, we provide tests for it in the present investigation. Nevertheless, in his review of self-efficacy research Pajares (2009) subsequently reiterated this distinction between self-concept and self-efficacy. Referring to the Marsh et al. (1991) study, Pajares (2009, p. 561) noted that:

By comparing one's own performance with those of others ("I am a better math student than most of my friends") and also one's own performance in related areas ("I am better at math than at English"), an individual develops a judgment of self-worth—a self-concept. Self-efficacy judgments, on the other hand, focus on the specific ability to accomplish the criterial task; hence, frame-of-reference effects do not play a prominent role.

Despite being highlighted by both self-concept and self-efficacy researchers, this distinction has not been emphasized in most discussion of the two constructs, and apparently has not been tested systematically in rigorous empirical research.

In summary, a critical, largely unexplored distinction between self-concept and self-efficacy measures is the extent to which they are influenced by negative frame-of-reference effects: negative, social comparison effects of school-average achievement (the BFLPE); negative, dimensional comparison effects of verbal achievement on math self-concept (the I/E model). In particular, we posit that these negative frame-of-reference effects should be largely or completely truncated in relatively pure measures of self-efficacy that "provide respondents with a specific description of the required referent against which to judge their competence" (Bong & Skaalvik, 2003, p. 9). Thus, if the appropriate frame of reference and context are fully contained in the self-efficacy item itself, then social and dimensional comparison effects should be substantially attenuated. Indeed, if all preexisting differences could be controlled, there should be no contextual effects, positive or negative, in self-efficacy responses that are purely descriptive.

Appropriate construction of self-efficacy items. It is important that empirical support for the aforementioned three distinctions between self-concept and self-efficacy depends on how measures of these constructs are constructed. Thus, comparing the self-concept and self-efficacy measures typically used in applied research (as opposed to relatively pure self-efficacy measures, consistent with the design features originally posited by Bandura and colleagues), Marsh et al. (1991; Marsh, 2007; Marsh, Trautwein, et al., 2008) noted that instruments claiming to measure self-efficacy are sometimes based on items that are likely to invoke social comparisons with other students. Hence, the distinction between instruments purporting to measure self-concept and self-efficacy is likely to depend more on the nature and wording of the items than on the label assigned to the construct.

Consistent with lessons from jingle-jangle fallacies, some generalized self-efficacy measures are indistinguishable from self-concept measures. Thus, for example, Marsh, Trautwein, et al.

(2008; see also Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014) argued that the generalized self-efficacy items in PISA 2000 were more like self-concept items, in that the criterion of successful performance was not an explicit part of these items (hereafter we refer to this type of measure as generalized self-efficacy). It is for this reason that they found a negative effect of school-average ability (the big-fish-little-pond effect, BFLPE) for self-efficacy responses, albeit one that was somewhat smaller than for academic self-concept.

Similar concerns exist for Generalized Math Self-Efficacy, as considered here (see item wording in [online supplemental materials](#), Section 1) in that items such as "I am convinced that I can perform well in math homework and on math tests" do not specify a clear criterion of what it means to perform well, and students have to adopt some frame of reference to respond to the item: for example, with the performances of their classmates (social comparison) or, perhaps, their accomplishments in other school subjects (dimensional comparison). In this respect, they are more like typical self-concept items than self-efficacy items (see, e.g., [online supplemental materials](#), Section 1, for the self-concept items used in the present investigation).

It is also relevant that what we refer to as generalized self-efficacy in PISA 2000 was dropped and replaced with a more task-specific measure of self-efficacy in PISA 2003 (Lee, 2009; OECD, 2004). Noting that Betz and Hackett (1983) found that task-specific math self-efficacy was a better predictor of career choice than test performance, OECD developed a similar task-specific measure of functional math self-efficacy that was more closely aligned to the design features of the pure self-efficacy items outlined earlier, but also consistent with the PISA approach of assessing mathematical literacy in relation to real-world problems. On the self-efficacy scale used in PISA 2003 and subsequent PISA data collections, students reported their confidence in relation to functional mathematical tasks (e.g., using a train timetable; calculating the price of a product after a 30% discount; the number of tiles needed to cover a floor of certain dimensions; interpretation of graphs; reading a map) as well as solving math equations like those in traditional standardized math tests and in test-related self-efficacy measures. In a comparison of the self-efficacy measures in PISA 2000 and PISA 2003, Huang (2012) noted that self-concept was a better predictor of achievement in PISA 2000, but self-efficacy predicted achievement better for PISA 2003. The PISA 2003 measure of self-efficacy is closely related to the Betz and Hackett (1983, 1993) Mathematics Self-Efficacy Scale. Betz and Hackett based their measure on three components: solving math problems like those found on standardized achievement tests, functional mathematical competencies used in everyday life, and capability to perform in math classes requiring various degrees of math mastery. Thus, the PISA 2003 measure of self-efficacy is primarily related to the second component proposed by Betz and Hackett.

For present purposes, we distinguish generalized self-efficacy items (as in the PISA 2000 instrument and in generalized self-efficacy items that are more like self-concept items) from purer self-efficacy measures that are more consistent with the design principles originally proposed by Bandura (1997; also see Bong & Skaalvik, 2003; Schunk & Pajares, 2005). However, we also distinguish between *test-related self-efficacy*, in which students are asked to evaluate their ability to answer questions that are similar

to or the same as test items subsequently presented as part of a standardized achievement test, and *functional self-efficacy*, based on items like those in the functional self-efficacy component of the [Betz and Hackett \(1983\)](#) measure. Operationally, in terms of a particular study, this distinction is straightforward; the test-related self-efficacy items are based on items that subsequently appear on the standardized test, whereas the functional self-efficacy items are like those on the Betz and Hackett instrument, which are constructed independently of the standardized achievement test. In practice, however, this distinction is not so clear-cut, given that it is possible to include functional items on a standardized math test or to include test-like items in a set of functional self-efficacy items (e.g., $3x + 5 = 17$ and $2(x + 3) = (x + 3)(x - 3)$ used in PISA 2003).

Marsh and Pajares debate on relevance of content-specificity.

The issues of content specificity, appropriate construction of test-related self-efficacy items, and distinctions between self-concept and self-efficacy were the foci of a protracted dialogue between self-efficacy researcher Frank Pajares and self-concept researcher Herb Marsh that culminated in a jointly authored “constructive dialogue” ([Marsh et al., 1997](#)). All four authors noted that educational researchers assess self-efficacy by asking students to rate their capability to complete target tasks (i.e., math test items) and then testing their performance on the same or similar items. Pajares argued that using the same items maximized self-efficacy’s predictive power, whereas Marsh countered that using the same items positively biased estimates of correlations between self-efficacy and test performance. They agreed that their results (a reanalysis of results from earlier studies by Pajares and colleagues) showed that these positive biases did exist, but that they were not large. However, particularly Marsh emphasized that the content specificity of self-efficacy and parallel performance measures is a double-edged sword, in that the measures can be so narrowly defined as to have limited relevance for a broader range of criteria. In contrast, [Pajares \(1996\)](#) argued that more generalized measures of self-efficacy are good predictors of a broader array of outcomes (subsequent test scores, school grades, future aspirations, and choice behavior). [Marsh et al. \(1997\)](#) also emphasized that it might be possible to construct self-efficacy items that tap different task-specific skills within a specific domain. These suggestions relate to what we refer to here as generalized measures of self-efficacy and functional self-efficacy, as well as to test-related self-efficacy, which had been the initial focus of the Marsh-Pajares dialogue.

In a strategic compromise to maintain a constructive dialogue while still agreeing to disagree, the authors agreed on several directions for further research to address ongoing areas of concern ([Marsh, Roche, Pajares, and Miller, 1997](#), pp. 375–376):

- (a) to more fully delineate the apparently overlapping constructs such as self-efficacy and self-concept on grounds other than domain specificity (since either construct could, conceivably, be measured at any level of domain specificity);
- (b) to explore the evaluative component(s) of self-efficacy responses that seem to be important to the ability of self-efficacy beliefs to guide future behavior but seem to be attenuated in operationalizations of self-efficacy in much educational research;
- (c) to evaluate further the theoretical and practical implications of more generalized or global self-efficacy measures in relation to the

widely heralded concern ([Bandura, 1986](#), in press [subsequently published in 1997]; [Pajares, 1996](#)) that such measures transform self-efficacy beliefs into a generalized trait that is antithetical to social cognition theory; and

(d) to pursue the thorny problems of the direction of causality of self-efficacy and other constructs using approaches that have been the focus of much self-concept research (e.g., [Marsh, 1993](#)) within the context of longitudinal studies.

Issues such as these can be pursued appropriately in multiwave studies in which the same constructs (e.g., self-efficacy, self-concept, and other relevant constructs measured at different levels of generality and a variety of outcome measures) are each measured on different occasions and analyzed with SEM techniques such as those demonstrated here. Although this ambitious agenda may be beyond the scope of any one study, it is consistent with [Pajares’ \(1996\)](#) call for increased “intertheoretical cross talk” in which researchers with differing theoretical allegiances engage in collaborative research using various designs, statistical models, and construct operationalizations that are consistent with their construct’s theoretical home.

In many respects, we begin with the program of research proclaimed in this pivotal dialogue/debate between Marsh, Pajares, and colleagues. Indeed, in the present investigation we propose to pursue the ambitious agenda proposed by [Marsh et al. \(1997\)](#) in a single, large-scale study. More specifically, here we evaluate distinctions between self-concept measures, generalized self-efficacy measures (which are more like self-concept measures), and appropriately defined self-efficacy measures, in relation to factor structure, frame-of-reference effects, ability to predict future performance and choices, and jingle-jangle fallacies.

Outcome Expectancy in Expectancy-Value Theories and Control-Value Theory

Other psychological constructs have also been developed to assess self-beliefs that add even more complexity to the murky distinction between self-concept and self-efficacy. Of particular relevance to the present investigation, the construct of outcome expectancy has been important since early theoretical work by [Tolman \(1932\)](#); [Lewin, Dembo, Festinger, and Sears \(1944\)](#) and, subsequently, [Atkinson’s \(1964\)](#) model of achievement motivation (1964). Modern versions of expectancy-value theory ([Eccles, 2009](#); [Eccles & Wigfield, 2002](#); also see discussion of related control-value theory; [Pekrun, 2006](#)) have greatly expanded on this historical theoretical framework, incorporating a wide variety of psychosocial and sociocultural variables. Indeed, self-efficacy theory ([Maddux, 2009](#)) proposes that self-efficacy perceptions are independent of outcome expectancy. Of particular relevance, [Eccles \(1984, 1987\)](#) initially posited academic self-concept to be distinct from expectations of success: Whereas academic self-concepts were posited as domain-specific competence beliefs, expectations of success were operationalized as more narrowly defined task-specific expectations of the likelihood of success on an upcoming task.

In early versions of expectancy-value theory (EVT), [Eccles \(1987\)](#) distinguished between outcome expectancy as a more self-efficacy-like construct, and academic self-concept. However, based on subsequent empirical research, Eccles and colleagues found that the two

constructs were relatively indistinguishable (Eccles & Wigfield, 1995; Guo, Marsh, Parker, Morin, & Dicke, 2017; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006). Similarly, Schunk and Pajares (2005) noted that this conceptualization of expectancy in expectancy-value theory is similar to that used in self-efficacy research, but also emphasized that expectancy-value theorists have subsequently concluded that expectations of success and domain-specific self-concept are not empirically separable (Eccles & Wigfield, 2002; Wigfield & Eccles, 1992). Indeed, as emphasized by Williams and Rhodes (2016), appropriately constructed self-efficacy measures should theoretically be independent of outcome expectancy and should precede it in terms of causal ordering. Furthermore, Wigfield et al. (2006) emphasized that competence beliefs in EVT, as in self-concept research (e.g., Harter, 1998; Marsh, 1990), are defined in relation to how good one is at a particular activity, relative to other individuals—an approach that is different to that used in self-efficacy research. Indeed, many recent EVT studies have used academic self-concept responses to operationalize expectations of success (e.g., Eccles, 2009; Guo, Marsh, Morin, et al., 2015; Guo, Parker, et al., 2015; Fredricks & Eccles, 2002; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Nagengast, Marsh, et al., 2011; Trautwein, Marsh, et al., 2012). Harter (1986, 1998, 2012) also has focused on students' perceptions of their own competence. However, like Eccles and Wigfield (2002), Harter operationalized competence perceptions as self-concept responses. Similarly, related studies of control-value theory (Pekrun, 2006) often define expectancy and perceived control operationally in terms of academic self-concept (see Marsh, Pekrun, Murayama, Arens, et al., 2017; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017; Pekrun & Perry, 2014). In this respect,

self-beliefs and generalized outcome expectancies are seen as empirically indistinguishable in current versions of expectancy-value theory (Eccles & Wigfield, 2002) and control-value theory (Pekrun, 2006) and are operationalized as self-concept responses, as in research by Harter (2012), and Marsh (1990, 2007).

This reconceptualization of outcome expectancy from a relatively self-efficacy-like construct to a relatively more self-concept-like construct in expectancy-value theories and control-value theory is highly relevant to our discussion of the murky distinction between self-concept and self-efficacy. Consistently with this reconceptualization, we posit that outcome expectancy related to broader outcomes in the future will be distinct from pure measures of self-efficacy and relatively indistinguishable from self-concept. Furthermore, generalized outcome expectancy should be subject to similar negative frame-of-reference (social and dimensional comparison) effects as self-concept, whereas appropriately constructed self-efficacy measures should not. Hence, on the basis of this logic, generalized outcome expectancy and self-concept measures reflect a jangle fallacy, in which two scales with apparently dissimilar labels actually measure similar constructs. Here we make this issue explicit in the broader conceptualization of our construct validity approach to the distinction between self-efficacy and self-concept, and to jingle-jangle fallacies.

The Present Study: A Priori Research Hypotheses and Research Questions

Here we more fully investigate a priori predictions about the murky distinctions between self-concept, self-efficacy, and outcome expectancy over time (Figure 2); a substantive-methodological synergy

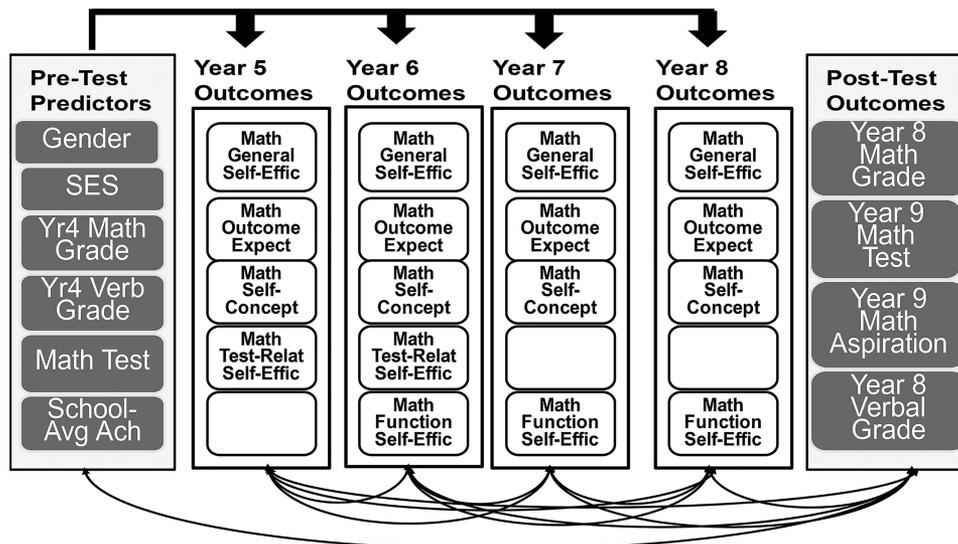


Figure 2. Schematic diagram of structural equation model. All 17 self-factors (4 or 5 factors for each of four waves of data) are latent variables, but individual items are not shown, to avoid clutter (see factor loadings in online supplemental materials, Supplemental Table 2). All pretest and posttest variables are single-item factors, with the exception of Math Aspirations, which is a latent factor. All 17 self-factors are regressed on the set of five pretest predictors (represented by the single-headed paths leading from the pretest predictors at the top of the diagram). All the remaining relations are represented as residual correlations, controlling for pretest variables (represented by the network of double-headed curved lines at the bottom of the diagram). Empty boxes reflect the fact that not all factors were collected at each wave of the study.

(Marsh & Hau, 2007), bringing together new substantive, theoretical and statistical models in a novel way not previously considered. The data set is based on a representative sample of 3,350 students from math classes in 43 German schools. These data included school grades in German and math from the year before the start of secondary school. Data (math school grades, standardized math achievement tests, and five self-belief measures) were then collected in all of the subsequent five years of compulsory secondary schooling. Posttest outcomes consisted of school grades, test scores, and future aspirations near the end of compulsory education, and were gathered after the final wave of self-belief measures (Figure 2) was collected. Consistent with the German school system, the primary schools considered here were not tracked in relation to achievement; schools and classes were relatively heterogeneous in relation to achievement in Year 4. However, from Year 5, primarily on the basis of Year 4 primary school performance, students in the Bavarian school system are typically tracked into three school types: high-achievement (*Gymnasium*), middle-achievement (*Realschule*), or low-achievement (*Hauptschule*) school tracks.

In summary, in our overall design (Figure 2) the main focus (self-belief outcome variables in Figure 2) is on five math self-belief measures collected during the first four years of secondary school (Years 5–8). However, we also consider math and German primary school grades from Year 4—the year prior to secondary school—gender, socioeconomic status (SES), and school-average achievement as predictor variables (Figure 2) to test frame-of-reference effects (BFLPE and the I/E model) that have been well validated in self-concept research. A novel contribution is the theoretical distinction between self-concept and self-efficacy in relation to these two frame-of-reference effects and on how well these frame-of-reference effects generalize to the other self-belief constructs considered here. Finally, we use posttest outcomes (math and German school grades, math test scores, future math aspirations) collected near the end of compulsory schooling to test how well self-belief constructs collected during each of the first four years of secondary school are able to predict these outcomes after controlling for pretest predictors (Figure 2). For present purposes we classify a priori the five math self-belief constructs (see online supplemental materials, Section 1, for the wording of the items) into two categories: three self-concept-like constructs (math self-concept, outcome expectancy, and generalized math self-efficacy) and two self-efficacy-like constructs (test-related math self-efficacy; functional math self-efficacy based on functional self-efficacy items from the Mathematics Self-Efficacy Scale developed by Betz & Hackett, 1983, 1993). In relation to results from an all-encompassing SEM represented in Figure 2, we offer the following research hypotheses and questions.

1. Latent correlations among five self-belief constructs over time
 - a. Self-concept-like factors: Within each wave of data (Years 5–8 in Figure 2), correlations among the three self-concept-like factors will be high enough to render them empirically indistinguishable. Across different waves of data, test–retest correlations among the three self-concept-like factors will provide good support for convergent validity in relation to stability over time, but little or no evidence of discriminant

validity in relation to these three self-concept-like constructs.

- b. Self-efficacy-like factors: The two self-efficacy-like factors will be distinct from the set of three self-concept-like factors within and across different waves. We leave as a research question whether the two self-efficacy-like factors are distinct from each other and, if so, whether each is more highly correlated with the other than with the self-concept-like factors.
2. Frame-of-reference effects (social and dimensional comparison effects)
 - a. Dimensional comparison effects (based on the I/E Model; Figure 1B). The negative effect of prior verbal achievement (German grades in Year 4) on the three math self-concept-like factors will be significantly negative across all four waves (i.e., in each of the Years 5–8; Figure 2); the corresponding effects on the two self-efficacy-like constructs will be substantially attenuated (i.e., considerably less negative or completely eliminated) compared with the negative effect of verbal achievement on the three self-concept-like constructs.
 - b. Social comparison effects (BFLPE; Figure 1A). The negative effect of school-average math achievement on the three math self-concept-like factors will be significantly negative across all four waves; the corresponding effects on the set of two self-efficacy-like constructs will be substantially attenuated (i.e., considerably less negative or completely eliminated).
 3. Long-term predictions based on self-efficacy and self-concept ratings
 - a. Based on a set of three math posttest outcomes (math school grades at the end of Year 8, math test scores in Year 9, and future math aspirations in Year 9), predictive relations (after controlling for pretest variables; Figure 2) will be higher for the math self-concept-like factors than for the math self-efficacy-like factors across each of the four waves. We leave as research questions whether these differences vary as a function of the posttest outcome, and how these math-self-beliefs are correlated with posttest verbal (German) school grades.
 4. Higher-order factor structure
 - a. Based on the hypothesis that the three self-concept-like factors measure essentially the same construct, we posit that within each wave they can be represented as a single higher-order factor with little loss in fit, and that support for Hypotheses 1–3 will be essentially the same when each of these first-order factors is treated as a separate factor.

Method

Participants and Sampling

Our study is a secondary data analysis based on the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Frenzel, Goetz, Lüdtke, Pekrun, & Sutton, 2009; Frenzel, Pekrun, Dicke, & Goetz, 2012; Marsh, Pekrun, Murayama, Guo, et al., 2017; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Murayama, Pekrun, Suzuki, Marsh, & Lichtenfeld, 2016; Pekrun et al., 2007, 2017). PALMA is a large-scale longitudinal study of the development of math achievement and related beliefs in the German federal state of Bavaria. It has 6 measurement waves (Years 5 to 10) in addition to school grades from the last year of primary school (Year 4).

The study used a stratified sampling of schools in the federal state of Bavaria, considering location (rural, urban, size of city, region within Bavaria), type of school (track), and school size. It is important to note that this is the same standard procedure as for the PISA assessments. Indeed, the sampling was carried out by the Data Processing and Research Center (DPC) of the International Association for the Evaluation of Educational Achievement (IEA), which also does the sampling for PISA in Germany. Thus, the sampling results in a student sample that was representative of Bavaria in terms of student characteristics such as gender, urban versus rural location, and SES (for details, see Pekrun et al., 2007).

For present purposes we focus on Years 5–8 and outcomes in Year 9, because mandatory education finishes after year 9 in Germany. Hence, data following Year 9 are no longer representative in regard to students' ability and achievement tracks. On the basis of primary school results, students ($N = 3,530$; 50% girls; mean age = 11.7 in Year 5, $SD = 0.7$) were allocated to the high-achievement (*Gymnasium*; 37%), middle-achievement (*Realschule*; 30%), or low-achievement (*Hauptschule*; 33%) school tracks.

Near the end of each successive school year, trained external test administrators administered the PALMA instruments. Participation in the study was voluntary. However, parental consent was obtained for all participating students, and the acceptance rate was a very high 91.8%; participation at the school level was 100%. Surveys were anonymized to ensure participant confidentiality.

Measures

Math achievement. Students' achievement was based both on school grades (end-of-the-year final grades obtained from school records) and on standardized achievement tests. Grades from Year 4 (the last year of primary school) were used as pretest covariates, and grades from Year 8 (at the end of the school year, after the Year 8 data collection) were used as posttest outcomes. Mathematics achievement was additionally assessed by the PALMA standardized math test (vom Hofe, Pekrun, Kleine, & Götz, 2002; vom Hofe, Kleine, Blum, & Pekrun, 2005; Murayama et al., 2013) on the basis of a combination of multiple-choice and open-ended items, using multimatrix sampling with a balanced incomplete block design (for details, see vom Hofe et al., 2002). The number and difficulty of items, varying between 60 and 90 items across the different waves, increased with each wave. Test scores for Years 5 and 9 were used as pretest predictor variables

and posttest outcomes, respectively. Year 5 test scores were also used to define class-average achievement for the purposes of testing the BFLPE.

Math self-belief measures. At each measurement wave students completed a detailed survey including the self-belief constructs (see online supplemental materials Section 1 for the wording of items, response scales and coefficient alpha estimates of reliability for each of the scales, and Section 2 for the factor loadings relating each item to its latent factor). For present purposes, as shown in Figure 2, these constructs were classified as either math-self-concept-like constructs or self-efficacy-like constructs, according to whether the items had a specific description of a referent against which to judge competence (see earlier discussion). The three math self-concept-like constructs were: self-concept (6 items; e.g., "In math, I am a talented student"); outcome-expectancy (6 items; e.g., "I am sure to get good marks in math exams when I try hard"); generalized self-efficacy (4 items; e.g., "I am convinced that I can perform well on math tasks and in math homework"). The two math self-efficacy-like constructs were: test-related self-efficacy (three items consisting of test items administered prior to the test in which students were asked "How confident are you that you can solve this math problem?") and functional self-efficacy (6 items based on the "using math in everyday tasks" from the Betz & Hackett, 1983, math self-efficacy scale; e.g., "How confident are you to be able to work out the price of a T-shirt when getting 20% off"). As shown in Figure 2, the three self-concept-like factors were measured in all four waves, but the two self-efficacy measures were only administered in the first two waves (test-related self-efficacy; Years 5 and 6) or the last three waves (functional self-efficacy; Years 6–8). A major focus of this study is to evaluate support for this a priori classification of self-belief constructs into these two categories, using a construct-validity approach in relation to jingle-jangle fallacies and a multitrait-multimethod analysis in relation to stability over time (Hypothesis 1).

Additional pretest predictors and posttest outcomes. The pretest predictors and posttest outcomes were based on math and German achievement measures (see earlier discussion and Figure 2). Additional pretest control variables consisted of students' gender and SES. SES was assessed by parent report, using the Erikson Goldthorpe Portocarero (EGP) social class scheme (Erikson, Goldthorpe, & Portocarero, 1979). An additional posttest outcome collected in Year 9 (Figure 2), the final year of mandatory schooling, consisted of a four-item scale designed to measure professional math aspirations following secondary schooling (e.g., "As an adult, I would like to be involved in many projects that are related to math"; see online supplemental materials for the wording of items and factor loadings).

Statistical Analyses

All analyses were done with Mplus 7.3 (Muthén & Muthén, 2008–14) using the robust maximum likelihood estimator (MLR), which is robust against violations of normality assumptions. Because students were clustered within schools, the Mplus complex design was used to appropriately adjust *SEs*. As is typical in large longitudinal field studies over such an extended period, many students had missing data for at least one of the measurement waves, attributable primarily to absence, changing schools, or

having entered the study after Wave 1. The numbers of waves of data completed by students were: 1 (17.0%), 2 (27.1%), 3 (10.8%), 4 (45.1%).

Particularly in longitudinal studies, there is increasing awareness of the limitations of traditional approaches to missing data (Graham, 2009). Here, to make full use of the data from students with missing data, we applied the full information maximum likelihood method (FIML; Enders, 2010). FIML has been found to result in trustworthy, unbiased estimates for missing values even in the case of large numbers of missing values (Enders, 2010) and to be an adequate method to manage missing data in large longitudinal studies (Jeličić, Phelps, & Lerner, 2009). More specifically, as emphasized in classic discussions of missing data (e.g., Newman, 2014), under the missing-at-random (MAR) assumption that is the basis of FIML, missingness is allowed to be conditional on all variables included in the analyses but does not depend on the values of variables that are missing. In a longitudinal panel design, this implies that missing values can be conditional on the values of the same variable collected in a different wave. This makes it unlikely that MAR assumptions are seriously violated, as the key situation of not MAR is when missingness is related to the variable itself. Hence, having so many waves of parallel data provides strong protection against this violation of the MAR assumption (see online supplemental materials, Section 3 for further discussion).

MTMM and Multitrait-multi-time-point (MTMTP) analyses. Campbell and Fiske's (1959) MTMM paradigm is, perhaps, the most widely used construct validation design to assess convergent and discriminant validity and is a standard criterion for evaluating psychological instruments, particularly in relation to self-concept measures (e.g., Byrne, 1996; Marsh & Hattie, 1996; Shavelson et al., 1976; Wylie, 1989). The rationale underlying MTMM designs is also ideal for evaluating jingle-jangle fallacies. Although the original Campbell-Fiske guidelines continue to be used widely, important problems with them are well known (e.g., Marsh, 1988, 1995; Marsh & Grayson, 1995). However, many subsequent CFA approaches to the evaluation of MTMM data are based on a single scale score—often an average of multiple items—to represent each trait-method combination. As argued by Marsh, Ellis, Parada, Richards, and Heubeck (2005; Marsh & Hocevar, 1988), multiple indicators for each trait-method combination should be incorporated into the MTMM analysis. Indeed, if this is done, confirmatory factor analyses at the item level results in an MTMM matrix of latent correlations, thereby eliminating most objections to the Campbell-Fiske guidelines.

In MTMM designs, the multiple methods traditionally refer to distinct ways of measuring the same constructs (e.g., ratings by self vs. others, or different approaches to measurement of the same set of constructs). Campbell and O'Connell (1967) subsequently proposed that multiple occasions in longitudinal data could serve as the multiple methods in their MTMM paradigm. To clarify this distinction, hereafter we use the expression MTMTP when referring to mono-method longitudinal studies in which time (i.e., the multiple time points) is treated as a method factor, as proposed by Campbell and O'Connell in their extension of the traditional MTMM design. Marsh et al. (2005; Marsh, Martin, & Jackson, 2010) also recommended this approach to evaluate convergent and discriminant validity in relation to temporal stability over time. When multiple time points are used as the method in MTMTP

designs, convergent validities refer to test-retest stability coefficients. Support for discriminant validity requires that correlations among different constructs on the same occasion (heterotrait-monotrait correlations in MTMM terminology) and correlations among different constructs on different occasions (heterotrait-heterotrait correlations) are smaller than convergent validities. Here we apply the logic of MTMTP to test the discriminant validity of the three self-concept-like factors in relation to each other and in relation to the two self-efficacy-like constructs (Hypothesis 1). More specifically, we posit that the three self-concept-like factors have little or no discriminant validity in relation to each other but do have discriminant validity in relation to the two self-efficacy-like factors.

Preliminary analyses: Factor structure. Analyses here are based on a single SEM, following from the design outlined in Figure 2. Our main focus is on correlations among the self-belief factors (Hypothesis 1), path coefficients relating the pretest predictor variables to the self-belief outcomes (Hypothesis 2), and the relations between the self-belief factors and the posttest outcomes (Hypothesis 3). However, it is important to emphasize that the model fit is good in relation to traditional indices (root mean square error of approximation = .019, comparative fit index = .944, Tucker-Lewis index = .937; see online supplemental materials Sections 2 and 3 for more details), and that the latent factors were well defined and consistent across the four waves of data (see supplemental materials, Supplemental Table 2 for factor loadings relating responses on 87 items to 17 latent self-belief and one posttest factor, and Section 3 for the Mplus syntax).

Results

Hypothesis 1: Latent Correlations Among Five Self-Belief Constructs Over Time (Table 1)

Latent correlations among the five self-belief factors over the four waves (Table 1) were used to test a priori predictions in Hypothesis 1. In evaluating these predictions, we adapt the logic of MTMTP analyses, in which the multiple occasions are seen as multiple methods. From this perspective, the test-retest correlations are evidence of convergent validity (in relation to stability over time), and the size of correlations among different constructs is used to infer discriminant validity. Thus, there is support for discriminant validity if test-retest correlations (stability of the same construct over time; correlations in bold in Table 1) are systematically larger than within-wave correlations among different constructs measured in the same wave (triangular subblocks outlined in bold black borders in Table 1) and between-wave correlations among different constructs (square subblocks in bold gray borders in Table 1).

Correlations among the three self-concept-like factors (Hypothesis 1a). Of particular relevance to Hypothesis 1a are the extremely large correlations among the three self-concept-like factors within each of the four waves of data (shaded in gray in Table 1). These 12 latent correlations vary from .88 to .97.

The test-retest correlations among the three factors were substantial, particularly for adjacent waves but also for correlations between responses in Years 5 and 8. However, test-retest correlations relating different constructs (among the three self-concept-

Table 1
Latent Correlations Among All Self-Belief Factors

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Year 5																	
1 MGSEff-Yr5	1.00																
2 MOutEx-Yr5	.91	1.00															
3 MSC-Yr5	.94	.87	1.00														
4 MTSEff-Yr5	.26	.25	.27	1.00													
Year 6																	
5 MGSEff-YR6	.57	.55	.58	.16	1.00												
6 MOUTEx-YR6	.50	.56	.52	.14	.92	1.00											
7 MTSEff-YR6	.58	.55	.66	.20	.94	.88	1.00										
8 MPSEff-YR6	.38	.33	.43	.37	.48	.41	.50	1.00									
9 MASEff-YR6	.41	.40	.44	.30	.63	.60	.62	.58	1.00								
Year 7																	
10 MGSEff-YR7	.48	.45	.51	.12	.63	.59	.64	.37	.43	1.00							
11 MOutEx-YR7	.40	.44	.45	.07	.55	.60	.58	.27	.36	.94	1.00						
12 MSC-YR7	.48	.43	.55	.13	.61	.57	.68	.36	.41	.96	.92	1.00					
13 MFSEff-YR7	.41	.36	.43	.26	.49	.47	.53	.51	.62	.62	.58	.63	1.00				
Year 8																	
14 MGSEff-YR8	.48	.42	.49	.13	.55	.50	.57	.34	.36	.70	.64	.70	.47	1.00			
15 MOutEx-YR8	.42	.40	.43	.07	.48	.49	.49	.28	.29	.62	.64	.63	.42	.95	1.00		
16 MSC-YR8	.48	.43	.52	.14	.53	.49	.60	.35	.36	.69	.64	.74	.48	.97	.92	1.00	
17 MFSEff-YR8	.36	.34	.41	.25	.40	.37	.43	.49	.53	.42	.40	.46	.61	.57	.55	.58	1.00

Note. MGSEff = Math Generalized Self-Efficacy; MOutEx = Math Outcome Expectancy; MSC = Math Self-Concept; MTSEff = Math Test-Related Self-Efficacy; MFSEff = Math Functional Self-Efficacy; Yr4–Yr9 refers to school years where Yr4 is the last year of primary school and Yrs 5–8 are the first five years of secondary school. Triangular blocks (with bold black borders) are within-wave correlations among different constructs collected in the same wave; shaded are correlations among the three self-concept-like factors hypothesized to lack discriminant validity (see Hypothesis 1). Square blocks (with bold grey borders) are between-wave constructs among different constructs and test-retest stability coefficients (along the diagonal, in bold). All $|r| \geq .06$ (in absolute value) are statistically significant ($p < .05$).

like factors) were typically as high as or higher than test-retest correlations relating the same constructs. Thus, for example, the test-retest stability for generalized self-efficacy over Years 4 and 5 was .57, but the correlations between this construct and self-concept (.58) were essentially the same.

In summary, there is good support for the classification of these three constructs as self-concept-like factors, and for Hypothesis 1a. The self-efficacy label given to generalized self-efficacy apparently represents a jingle fallacy (in that it measures a construct that is different from self-efficacy) and a jangle fallacy (in that it is more appropriately seen as a measure of self-concept than self-efficacy). Also, consistent with the jangle fallacy is that even though the math self-concept and outcome expectancies were given different labels, they apparently reflect a similar construct.

Correlations involving the two self-efficacy-like factors (Hypothesis 1b). Test-related self-efficacy was only measured in Years 5 and 6, while functional self-efficacy was measured in Years 6, 7, and 8 (Table 1 and Figure 2). Nevertheless, in support of Hypothesis 1b, within each of the four waves there was evidence that these two factors were relatively distinct from the three self-concept-like factors (correlations of .25 to .63), particularly in relation to the extremely high correlations among the self-concept-like factors (.88 to .97).

For test-related self-efficacy, the one test-retest stability coefficient (self-efficacy in Years 5 and 6) was only .37. Indeed, test-

related self-efficacy in Year 6 was as highly correlated with the three Year 6 self-concept-like factors (.33 to .43) as test-related self-efficacy in Year 5 (.36). Hence, from the perspective of MTMTP, support is not particularly strong even for convergent validity (stability over time), and support for discriminant validity is weak—at least in comparison to convergent validity.

For the functional self-efficacy measure, there were three test-retest stability coefficients (.53–.62). Thus, functional self-efficacy was clearly more stable than test-related self-efficacy. Furthermore, these stability coefficients were consistently higher than correlations between functional self-efficacy and the three self-concept-like factors. Thus, for example, Year 8 functional self-efficacy correlated .61 with Year 7 functional self-efficacy, but only .40 to .46 with the three self-concept-like measures. Hence, there is reasonable support for functional self-efficacy in relation to convergent validity and discriminant validity on the three self-concept measures.

Only in Year 6 were both self-efficacy measures collected. The correlation between them was only .58—marginally lower than correlations of functional self-efficacy with the three self-concept-like factors (.60 – .63)—and marginally higher than correlations of test-related self-efficacy with the three self-concept-like factors (.41–.49). Hence, the two self-efficacy measures (unlike the three self-concept-like factors) were not so highly correlated with each other so as to be considered the same construct.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Hypothesis 2: Frame-of-Reference Effects (Social and Dimensional Comparison Effects; Table 2)

Dimensional comparison (I/E) effects. The I/E model predicts that paths from math achievement to math self-beliefs are positive; not surprisingly, the paths are consistently positive for all five self-belief constructs. However, the critical prediction for the I/E model is that the paths from verbal achievement (German grades in Year 4) to the math self-belief factors are negative. We hypothesized that these negative paths (and support for the I/E model) would be much stronger for the three math self-concept-like factors, and substantially (or completely) attenuated for the two self-efficacy-like factors.

Consistently with Hypothesis 2a, all 12 paths leading from verbal achievement to the three math self-concept-like factors were significantly negative ($-.13$ to $-.24$). In marked contrast, all five paths leading from verbal achievement to the two self-efficacy-like factors were close to zero and nonsignificant (.00

to $-.08$). Thus, as hypothesized, there was clear support for dimensional comparison effects as predicted by the I/E model for the self-concept-like factors, but not for the self-efficacy-like factors. Whereas support for the I/E model was marginally stronger for the math self-concept measure than for the other two math-self-concept-like factors (generalized self-efficacy and outcome expectancy), these differences were small, and support for predictions generalized across all three math-self-concept-like factors.

Social comparison (BFLPE) effects. The BFLPE model predicts that paths from math achievement to math self-beliefs are positive (as already shown for the I/E model). However, the critical prediction for the BFLPE is that the paths from class-average math achievement (L2Ach in Table 2) are negative. We hypothesized that negative paths (and support for the BFLPE) would be much more negative for the math self-concept-like factors, and substantially (or completely) attenuated for the self-efficacy-like factors.

Consistently with Hypothesis 2b, all 12 paths leading from class-average achievement to the three self-concept-like factors were significantly negative ($-.23$ to $-.30$). In marked contrast, none of the five paths leading from class-average achievement to the two self-efficacy-like factors was significantly negative ($-.02$ to $+.19$) and one was significantly positive ($+.19$). Thus, as hypothesized, there is clear support for the BFLPE for the self-concept-like factors but not for the self-efficacy-like factors. Whereas support for the BFLPE was marginally stronger for the math self-concept measure than for the other two math-self-concept-like factors (generalized self-efficacy and outcome expectancy), these differences were again small (e.g., BFLPE was $-.26$, $-.21$, and $-.22$ for self-concept in Year 8), and support for predictions generalized across all three math-self-concept-like factors and all 4 years (Table 2).

Summary of frame-of-reference effects. Across the two frame-of-reference effects (I/E and BFLPE model), four waves of data, and three self-concept-like factors, we predicted that all 24 paths representing these frame-of-reference effects would be negative; indeed, all 24 were significantly negative. Similarly, we predicted that all 10 paths representing these frame-of-reference effects for self-efficacy-like factors would not be substantially negative; 9 of the 10 were nonsignificant and one was significantly positive rather than negative. Hence, consistently with a priori predictions, the frame-of-reference effects that have been such an important feature of recent studies of self-concept formation are completely absent in the two self-efficacy-like measures.

In nearly all longitudinal studies the size of relations between predictors and outcomes diminishes over time, as was the case in our study for relations between self-belief measures and posttest outcomes. However, for the dimensional comparison (I/E) and social comparison (BFLPE) effects, the negative frame-of-reference effects actually increased from Year 5 to Years 6 and 7 and declined only slightly in Year 8. Hence, the frame-of-reference effects were very robust over time. In juxtaposition to these robust negative frame-of-reference effects for self-concept-like factors, those for self-efficacy-like factors were either nonsignificant or positive.

Table 2
Path Coefficients Regressing Self-Factors on Six Pretest Predictors on Four Occasions (Years 5–9)

Pretest predictors	Self-belief factors (outcome variables)									
	MGSEff		MOutEx		MSC		MTSEff		MFSEff	
	Path	SE	Path	SE	Path	SE	Path	SE	Path	SE
Year 5										
L2ACH	-.23	.05	-.23	.05	-.26	.05	.19	.05		
MTest-Yr5	.37	.03	.36	.04	.43	.03	.23	.05		
MGrd-YR4	.16	.04	.16	.05	.27	.04	.08	.05		
VGrd-YR4	-.17	.05	-.17	.05	-.24	.04	.00	.04		
SES	.05	.02	.03	.02	.04	.02	-.01	.03		
Boy	.15	.03	.06	.03	.15	.03	.04	.03		
Year 6										
L2ACH	-.30	.06	-.28	.05	-.32	.06	.14	.08	.03	.06
MTest-Yr5	.33	.04	.30	.04	.42	.03	.28	.05	.29	.03
MGrd-Yr4	.27	.04	.24	.04	.29	.04	.21	.06	.16	.04
VGrd-Yr4	-.22	.04	-.19	.05	-.24	.04	-.08	.05	-.04	.04
SES	.05	.02	.05	.02	.02	.02	-.01	.03	.05	.02
Boy	.13	.03	.03	.03	.11	.03	.15	.04	.12	.03
Year 7										
L2ACH	-.28	.05	-.28	.05	-.30	.05			-.02	.05
MTest-Yr5	.32	.03	.30	.04	.36	.03			.36	.04
MGrd-Yr4	.26	.04	.20	.04	.28	.04			.15	.05
VGrd-Yr4	-.17	.05	-.14	.04	-.19	.04			-.05	.04
SES	.01	.03	-.00	.03	-.01	.03			-.00	.03
Boy	.14	.02	.08	.02	.15	.02			.14	.03
Year 8										
L2ACH	-.22	.04	-.21	.04	-.26	.04			.00	.05
MTest-Yr5	.28	.04	.27	.04	.32	.04			.30	.04
MGrd-Yr4	.22	.06	.17	.04	.25	.05			.14	.05
VGrd-Yr4	-.14	.06	-.13	.05	-.16	.05			-.02	.05
SES	.03	.03	.03	.02	.02	.03			.01	.03
Boy	.14	.02	.08	.03	.15	.02			.13	.02

Note. Path = standardized path coefficient; SE = standard error; all coefficients in bold are statistically significant ($p < .05$); MGSEff = Math Generalized Self-Efficacy; MOutEx = Math Outcome Expectancy; MTSEff = Math Test-Related Self-Efficacy; MFSEff = Math Functional Self-Efficacy; Yr4–Yr9 year in school. MGrd = math grade (mark); VGrd = verbal (German, the native language) grade (mark). MTest = standardized math achievement test; L2Ach = Class-average achievement based on the MTest-Yr5. All 17 self factors were regressed on the set of 6 pretest covariates, whereas all other relations were represented by correlations (Figure 2).

Hypothesis 3: Long-Term Predictions Based on Self-Efficacy and Self-Concept

Our main focus here is on the relations between the five math self-belief factors in Years 5–8 and the three math posttest outcomes (math grades, math test scores and math future aspirations; Table 3). We predicted that after controlling for pretest predictors (Figure 2), these relations would be consistently more positive for the self-concept-like factors than for the self-efficacy-like factors. Consistently with predictions, within each wave of data the average correlation between the three self-concept-like factors and three posttest math outcomes was consistently higher than the corresponding average correlation between the self-efficacy-like factors and the posttest math outcomes.

As anticipated (but left as a research question), the results varied as a function of the particular wave of data, the particular self-belief construct, and particular posttest outcome. Not surprisingly, the correlations between self-belief factors (both self-concept and self-efficacy) increased as their measurement became more temporally proximal to the posttest outcomes (i.e., correlations with posttest outcomes were higher for Year 8 self-beliefs than for Year 5 self-beliefs). Thus, for example, math aspirations (in Year 9) correlated .42 with Year 8 math self-concept responses, but .35, .28, and .24 with math self-concepts measured in Years 7, 6, and 5. Nevertheless, the stronger relations for self-concept-like factors compared with self-efficacy-like factors were evident in each of the four waves of self-belief measures.

The differences between correlations for self-concept-like and self-efficacy-like variables did vary across the three math posttest

outcomes. The differences were consistently large for math grades and math future aspirations, but smaller for math test scores. This relatively better performance of self-efficacy-like variables in predicting standardized math test scores is not surprising, given that the test-related self-efficacy items in particular were based on math test items. However, it is interesting to note that on the basis of Year 6 measures (the only year in which both self-efficacy-like measures were collected), the correlations between the two self-efficacy-like measures and the three posttest math outcomes are very similar. In particular, both the test-related self-efficacy and functional self-efficacy scores had similarly small, but statistically significant correlations with the posttest standardized achievement scores (.06 and .05, respectively); the corresponding correlations with the three self-concept-like measures (.10 to .12) were also statistically significant and small, but somewhat larger. Indeed, within each of the four waves across all three posttest math outcomes, the highest relation was consistently with math self-concept.

Correlations between math self-belief factors and posttest German grades were smaller than for the math posttest outcomes. Math self-beliefs in Years 5 and 6 had small and mostly nonsignificant correlations with posttest German grades (–.07 to +.06). Even in Years 7 and 8 the correlations were not large (.04 to .14), although most were statistically significant. These results are consistent with the domain specificity of academic self-beliefs and, in this respect, support their discriminant validity. However, the difference in correlations with math and German grades was consistently larger for the self-concept-like factors than for the

Table 3

Correlations Relating Self-Belief Factors From Each Year to the Set of Four Posttest Outcomes (Controlling for the Effects of Pretest Covariates (Table 2))

Math outcomes	MGSEff With		MOutEx With		MSC With		MTSEff With		MFSEff With	
	Corr	SE	Corr	SE	Corr	SE	Corr	SE	Corr	SE
Year 5										
MGrdYr8	.13	.03	.12	.03	.17	.03	–.01	.04		
MTestYr9	.06	.02	.06	.02	.07	.02	.03	.02		
MAspireYr9	.21	.03	.15	.04	.24	.03	.01	.03		
Year 6										
MGrdYr8	.14	.03	.14	.03	.20	.03	.06	.04	.04	.03
MTestYr9	.10	.02	.11	.02	.12	.02	.06	.03	.05	.02
MAspireYr9	.24	.03	.21	.02	.28	.02	.15	.03	.16	.03
Year 7										
MGrdYr8	.29	.02	.30	.03	.34	.02			.09	.03
MTestYr9	.10	.02	.12	.02	.14	.02			.08	.02
MAspireYr9	.31	.03	.27	.03	.35	.02			.17	.03
Year 8										
MGrdYr8	.46	.02	.46	.02	.52	.02			.14	.03
MTestYr9	.15	.02	.16	.02	.17	.02			.16	.02
MAspireYr9	.39	.02	.34	.03	.42	.02			.27	.03
Verbal outcomes										
Year 5	.00	.02	–.01	.03	.01	.03	–.06	.04		
Year 6	.01	.02	.05	.02	.03	.02	.01	.03	–.01	.03
Year 7	.07	.03	.09	.03	.09	.03			.04	.03
Year 8	.10	.02	.13	.02	.11	.03			.08	.02

Note. Path = standardized path coefficient; SE = standard error; all coefficients in bold are statistically significant ($p < .05$); MGSEff = Math Generalized Self-Efficacy; MOutEx = Math Outcome Expectancy; MTSEff = Math Test-Related Self-Efficacy; MFSEff = Math Functional Self-Efficacy; Yr4–Yr9 year in to school. MGrd = math grade (mark); VGrd = verbal (German, the native language) grade (mark); MTest = standardized math achievement test; MAspireYr9 = Aspirations to study and use math in future. All coefficients are residual correlations after controlling for the set of six pretest covariates (Figure 2).

self-efficacy-like factors. Thus, for example, in Year 8 math self-concept-like factors were much more highly correlated with math grades (.46 to .52) than with German grades (.11 to .14), whereas for math functional self-efficacy the sizes of the differences were smaller (.15 vs. .08). From this perspective, even these results support the greater domain specificity of the self-concept-like factors than the self-efficacy-like factors.

Hypothesis 4: Higher-Order Factor Structure Combining the Three Self-Concept-Like Factors into a Single Higher-Order Factor

Based on the prediction that the three self-concept-like factors measure essentially the same construct, we hypothesized (Hypothesis 4) that these three first-order factors could be represented by a single higher-order factor (HO-math self-concept). In a preliminary evaluation of this solution, the fit was very good and differed little from the corresponding first-order solution (Table 4 and

online supplemental materials, Section 2). The higher-order solution was much more parsimonious, requiring 150 fewer parameter estimates. In particular, relations between the three self-concept factors and each of the other variables were represented by a single estimate rather than three separate estimates. The standardized factor loadings relating the first-order factor to the math HO-self-concept varied from .92–.99 (see Table 4). Given the excellent fit of the model, it is not surprising that support for each of the first three hypotheses was replicated with this higher-order structure.

Thus, for example, each of the test-retest correlations for the HO-math self-concept factor (Years 5–8) was approximately the average of the test-test correlations for each of the three self-concept-like factors considered separately. With the reduced number of factors, the correlations show more clearly that the self-efficacy-like factors are distinct from the HO-math self-concept factor.

Table 4

Results Based on Higher-Order Structural Equation Model Combining the Three Math-Self-Concept-Like Factors Into a Single Higher-Order Construct

Variable	Higher-order math self-concept				Test-related self-efficacy		Functional self-efficacy		
	Year 5	Year 6	Year 7	Year 8	Year 5	Year 6	Year 6	Year 7	Year 8
Higher-order factor loadings ^a									
MGSEff	.98	.98	.99	.99					
MOutEx	.92	.92	.94	.95					
MSC	.96	.98	.98	.98					
HO-MS-C-Yr5	1.00								
HO-MS-C-Yr6	.61	1.00							
HO-MS-C-Yr7	.50	.64	1.00						
HO-MS-C-Yr8	.49	.56	.70	1.00					
MTSEff-Yr5	.27	.17	.12	.11		1.00			
MTSEff-Yr6	.41	.48	.35	.33		.38	1.00		
MFSEff-YR6	.44	.64	.41	.36	.30	.58	1.00		
MFSEff-YR7	.42	.52	.63	.48	.26	.51	.62	1.00	
MFSEff-YR8	.39	.42	.45	.58	.25	.49	.53	.61	1.00
Path coefficients regressing self-factors on pretest predictors (Table 2) ^b									
L2ACH	–.30	–.33	–.29	–.27	.16	.08	–.02	–.03	–.02
MTest-Yr5	.39	.39	.35	.32	.22	.29	.29	.36	.31
MGrd-Yr4	.27	.27	.24	.22	.09	.23	.16	.14	.14
VGrd-Yr4	–.23	–.23	–.19	–.15	.03	–.05	–.03	–.05	–.02
SES	–.04	–.04	.00	–.03	.01	.01	–.04	.00	–.01
Boy	.11	.11	.14	.14	.04	.15	.12	.14	.13
Correlations relating self-belief factors to posttest outcomes (Table 3) ^c									
MAspireYr9	.21	.25	.33	.40	.00	.14	.17	.17	.27
MTestYr9	.07	.11	.13	.16	.06	.07	.06	.08	.16
MGrdYr8	.14	.16	.32	.50	–.03	.04	.04	.09	.14
VGrdYr8	.00	.03	.09	.12	–.07	.00	–.01	.04	.08

Note. HO-MS-C = Higher-order math self-concept factor (combining 3 self-concept-like factors); MTSEff = Math Test-Related Self-Efficacy; MFSEff = Math Functional Self-Efficacy; MGrd = math grade (mark); VGrd = verbal grade (German, the native language); MTest = standardized math achievement test; L2Ach = class-average achievement based on the MTest-Yr5; MAspireYr9 = Aspirations to study and use math in future. Yr4–Yr9 refer to school years where Yr4 is the last year of primary school and Yrs 5–8 are the first four years of secondary school. All coefficients in bold are statistically significant ($p < .05$), those not bolded and in italics are not statistically significant. The fit of this model was very good (RMSEA = .019, CFI = .940, TLI = .935) and differed little from the corresponding first-order factor structure (RMSEA = .019, CFI = .944, TLI = .937; see online supplemental materials, Supplemental Table 1, for more detail).

^a All coefficients are standardized factor loadings relating each of three self-concept-like first-order factors to the higher-order factor (higher-order math self-concept) separately for each wave. ^b All self factors were regressed on the set of six pretest covariates (Figure 2). Coefficients are path coefficients in the form of standardized regression coefficients. ^c All coefficients are residual correlations after controlling for the set of six pretest covariates (Figure 2).

The frame-of-reference effects based on the HO-factor structure are highly similar to those for the single factors (Table 2). In support of Hypothesis 2a (dimensional comparison effects based on the I/E model), the effect of verbal achievement on the HO-self-concept factor is significantly negative for all four years ($-.15$ to $-.22$); the corresponding effects for the self-efficacy-like factors were all nonsignificant. In support of Hypothesis 2b (social comparison effects based on the BFLPE), the effect of class-average achievement on HO-math self-concept was significantly negative for all four years ($-.24$ to $-.31$); the corresponding effects for the self-efficacy-like factors were not significantly negative in any of the 4 years and were significantly positive in one instance.

In support of Hypothesis 4, relations between the HO-math self-concept and the three posttest math outcomes (math future aspirations, grades, and test scores) were also similar to those in Table 3. In particular, the three criteria were consistently more highly correlated with HO-math self-concept than any of the self-efficacy-like factors within each of the four waves of data. In summary, in support of Hypothesis 4 and the supposition that the three self-concept-like factors are indistinguishable, the fit of the higher-order factor structure is good, and the results replicate the frame-of-reference effects and relations with posttest outcomes found for each of the first-order self-concept-like factors.

Discussion

The present investigation represents one of the most comprehensive studies ever undertaken into the murky distinction between self-concept and self-efficacy, offering new insights to this frequently discussed issue. Highlighting the importance of considering potential jingle-jangle fallacies, we argue that three self-belief constructs with deceptively distinct labels (generalized self-efficacy, outcome expectancies, and self-concept) actually measured very similar constructs and indeed, were essentially indistinguishable (correlations mostly greater than .9). This was particularly important in that one of these constructs (generalized self-efficacy) actually purported to be a self-efficacy measure, whereas historically, outcome expectancies have been seen to be more closely related to self-efficacy than to self-concept. Also, the finding that self-concept and outcome expectancies are nearly indistinguishable is consistent with earlier discussion and the fact that expectancy-value researchers are frequently using self-concept to operationalize the expectancy construct. From this perspective, there was good support for our classification of these three as self-concept-like measures (see Eccles & Wigfield, 2002; Wigfield, Tonks, & Klauda, 2016, for reviews).

The two self-efficacy-like constructs were clearly more distinct from each other ($r = .58$, Table 1). Indeed, the correlation between the two self-efficacy measures was sufficiently low that they were apparently measuring somewhat different constructs—different from self-concept, but also different from each other. Indeed, although there was support for the discriminant validity of the functional self-efficacy measure from the three self-concept measures, functional self-efficacy tended to be more highly correlated with the self-concept-like measures than with test-related self-efficacy.

On the basis of our brief review of relevant literature, three key distinctions between self-concept and self-efficacy are particularly

relevant. These distinctions relate to frame-of-reference effects (Hypothesis 2) and prediction of future accomplishments (Hypothesis 3).

In regard to Hypothesis 2, appropriately constructed self-efficacy items are designed to be purely descriptive, whereas self-concept items are both descriptive and evaluative. In particular, appropriately constructed self-efficacy measures provide respondents with the referent standard (frame-of-reference) as part of the item. In this respect they are specifically designed to substantially attenuate frame-of-reference (social and dimensional comparison) effects, which are so important in the formation of self-concept. Thus, self-efficacy researchers argue that self-efficacy is more purely descriptive of expected accomplishments but that self-concept is based on how accomplishments meet standards of worthiness associated with various frames of reference. We found remarkably strong support for our hypothesis that negative effects of verbal achievement (I/E model) and of school-average achievement (BFLPE) would be strong for self-concept-like factors, but substantially attenuated in relation to self-efficacy responses; all 24 frame-of-reference effects were significantly negative across the three self-concept measures and four waves of data, but none of the corresponding 10 frame-of-reference effects for self-efficacy-like factors was significantly negative.

With respect to Hypothesis 3, self-efficacy is future oriented (what can I do) whereas self-concept is based on past accomplishments. Superficially, this would seem to imply that self-efficacy should be more strongly related to future choices and accomplishments. However, there are several reasons why this might not be the case. In particular, Marsh (2007; Marsh et al., 1997) argued that because self-efficacy responses are designed to attenuate the value component and worthiness in self-descriptions of what one can do, this feature is also likely to attenuate their ability to predict future criteria that are based at least in part on self-evaluations of worthiness. Bong and Clark (1999) also acknowledge that self-concept responses are likely to be more inclusive than self-efficacy measures, reflecting evaluative inferences as well as more purely descriptive beliefs of what one can do. On this basis, we hypothesized that after controlling for pretest covariates, self-concept responses should be more highly correlated with posttest outcomes than are self-efficacy responses—particularly with those criteria that might involve a choice or motivation component. Again, support for this prediction was very strong, in that for each of the four waves of data, the three self-concept measures were more highly correlated with the math posttest outcomes (school grades, test scores, future aspirations) than with the self-efficacy measures. In line with the domain-specificity principle, the correlations between self-belief factors and posttest verbal achievement were consistently weak. However, even here the differences in correlations between math self-beliefs and math versus verbal posttest outcomes were systematically larger for the self-concept-like factors than for the self-efficacy-like factors.

Finally, consistently with Hypothesis 4, the higher-order factor structure (combining the three self-concept-like factors into a single higher-order factor) provided a well-defined factor structure that fitted the data well and replicated support for Hypotheses 1, 2, and 3.

Why Were Self-Efficacy Measures Not More Highly Correlated With Pre- and Posttest Variables?

The negative frame-of-reference effects observed for three self-concept-like factors (and HO-math self-concept) were noticeably absent for the two self-efficacy-like factors. This pattern of effects, of course, is consistent with the different rationales used to construct self-concept and self-efficacy items. In particular, because the standard of success is part of the construction of self-efficacy items, social and dimensional comparison effects are substantially attenuated. Indeed, for the results here, the negative frame-of-reference effects associated with both the I/E and the BFLPE model were completely eliminated. However, because of the removal of the “worthiness” or evaluative component from the self-efficacy responses, their ability to predict the critical posttest outcomes (school grades, test scores, and future aspirations, collected near the end of mandatory schooling in the German school system) was also attenuated, in keeping with suggestions by Marsh (2007) and with Hypothesis 3.

It is also important to emphasize that the longitudinal results presented here are addressing a different question than is typically addressed in cross-sectional—and even some longitudinal—studies comparing self-concept and self-efficacy. Thus, for example, the residual correlations (controlling for pretest predictors) relating self-belief factors to the posttest outcomes are clearly higher for the self-concept-like factors than for the self-efficacy-like factors (Table 3; also see Table 4). However, the zero-order correlations (without controls for pretest covariates) tend to be higher for self-efficacy-like factors than self-concept-like factors (see full set of correlations in [online supplemental materials](#), Supplemental Table 3). These results are easily explained in terms of the nature of the self-efficacy items. Particularly for the test-related self-efficacy measures, the content of the items closely parallels the actual content of the tests—indeed, items in test-related self-efficacy measures might actually be test items. Hence, it is not surprising that test-related self-efficacy scales were more highly correlated with test scores than were the self-concept-like factors.

However, the focus here is on the incrementally predictive power of the self-efficacy items after controlling for preexisting differences—including the standardized achievement tests that they are based upon. Although this rationale would seem to be less relevant to the functional self-efficacy factor, it is interesting to note that the residual correlations based on the two self-efficacy measures are nearly the same in Year 6 (see Table 2), the only year in which both self-efficacy-like measures were administered. We also note that different researchers have used the term incremental validity in different ways. Thus, for example, Huang (2012) focused on incremental validity in relation to another belief construct—self-concept or self-efficacy—to predict relevant criteria after controlling for the other. The diverging uses of the term is an important issue, but we caution that self-efficacy’s ability to predict outcomes is substantially attenuated when controlling for preexisting differences (including prior achievement), apparently to a much greater extent than is self-concept. Hence, we caution researchers to at least consider controls for pretest variables when comparing the predictive validity of self-concept and self-efficacy.

What Is the Role of Generalized Self-Efficacy Measures?

Self-efficacy researchers are facing a difficult challenge in establishing the generalizability of their measures. Historically, self-efficacy measures have been highly task-specific, focusing on extremely narrow content. This was appropriate so long as the focus of prediction was also similarly narrow. Such measures are potentially very useful in relation to a very narrow range of content. Generalizability might be argued in terms of the highly prescriptive manner in which self-efficacy items are constructed, but apparently not in relation to the generalizability of the content of such measures. This creates a quandary for test constructors who seek to develop standardized measures of self-efficacy that can be used in a range of studies and evaluated in relation to traditional psychometric criteria (i.e., factor structure, reliability, convergent and discriminant validity). Three quite different approaches to this problem have been taken.

Generalized self-efficacy measures. One strategy has been to develop generalized measures of self-efficacy in which the range of content is very general, rather than highly task-specific. The generalized math self-efficacy measure considered here (based on the General Self-Efficacy Scale; Schwarzer & Jerusalem, 1995) is a domain-specific version of this strategy, but many other generalized self-efficacy measures are designed to be domain general. Although the measures based on this approach might be useful, there are several grounds on which it can be argued that they are not really self-efficacy measures (see discussion by Maddux, 2009). Theoretically, they lack the domain or task specificity that was the original touchstone of self-efficacy measures. Also, as self-efficacy measures become more generalized, it becomes increasingly difficult to include concrete standards (task-specific frames-of-reference) as part of the item content. Without this design feature, the evaluative component of responses and the associated frame-of-reference effects that self-efficacy measures are designed to eliminate, are likely to play an increasingly important role—something that would seem to be anathema to self-efficacy theorists. Consistent with theoretical concerns, empirically such generalized measures of self-efficacy seem to measure a construct more closely related to self-concept than self-efficacy—as was the case in the present investigation.

In summary, the contention here is not that generalized self-efficacy measures are “bad” measures or that they are not potentially useful, but merely that they are not self-efficacy measures (at least, not pure measures of self-efficacy that are consistent with the original design features proposed by Bandura and other self-efficacy researchers—see earlier discussion). Support for this contention was very strong in the present investigation, and we suspect that similar issues may be evident in other generalized measures of self-efficacy. Thus, as noted by Pintrich (2003, p. 109), “at more global levels, self-efficacy beliefs would become more similar to perceived competence or self-concept, at least in terms of the motivational dynamics and functional relations to student outcomes.” Similarly, Pajares (2009) had a particular concern about researchers inappropriately labeling competence perceptions as self-efficacy perceptions. Indeed, it would seem that the very notion of a generalized self-efficacy measure is antithetical to the original underpinnings of self-efficacy research and theory. Clearly, for researchers who develop and use generalized measures

that are claimed to measure self-efficacy, the onus is on them to defend their measures in relation to theoretical, design, construct validity, and empirical considerations, such as those demonstrated here.

Functional self-efficacy measures. Functional self-efficacy measures like the [Betz and Hackett \(1983\)](#) measure used here, or the similar measure used by the OECD starting with PISA 2003, provide an alternative strategy to the development of generalizable measures of self-efficacy. In this approach, rather than relinquishing the task-specificity that is central to self-efficacy, researchers have developed multiple task-specific items within a given domain (e.g., math) that collectively cover a broader content range than is typically the case in self-efficacy measures. Although this strategy apparently avoids some of the pitfalls of generalized self-efficacy measures, there are still important measurement issues that have not been fully resolved, and that have important theoretical implications for self-efficacy research. In particular, the relevance of any particular set of items and the specific skills they target is likely to be highly dependent upon the cognitive developmental levels of respondents. Functional self-efficacy items developed for young children would not be appropriate for high school students, and those developed for high school students might not be appropriate for university students. Furthermore, to the extent that the nature of what is being measured by functional self-efficacy items is age-dependent, it might call into question the appropriateness of measuring change based on a fixed set of items in longitudinal studies. Issues such as these are not insurmountable; indeed, these are precisely the type of issues faced in the development of standardized achievement tests. Thus, longitudinal studies of standardized achievement routinely base estimates on large pools of different sets of items appropriate to the age and cognitive development of participants at each wave of data collection. Given appropriately chosen anchor items, sophisticated test-equating procedures are then used to place all the estimates on a common metric that generalizes across the different waves. We also note that a similar issue arises even in cross-sectional research covering a broad range of ages and years in school, such that the same functional self-efficacy item is likely to have different implications depending upon the relevant coursework that a given student has completed. Although this issue is particularly relevant to self-efficacy measures for students, similar problems exist in relation to other domains and age groups. Hence, although sophisticated methodological approaches to address these problems do exist, they apparently have not been applied to the development and testing of measures of functional self-efficacy.

Test-related self-efficacy. Here, we differentiated between measures of test-related self-efficacy and functional self-efficacy. Operationally, in terms of the present investigation, this distinction is straightforward; the test-related self-efficacy items were based on items that subsequently appeared on the standardized test, whereas the functional self-efficacy items were based on the existing [Betz and Hackett \(1983\)](#) math self-efficacy instrument. In practice, however, this distinction is not so straightforward, in that “functional” items might be included in standardized tests, and even items such as solving equations that typically appear on standardized tests could be included as functional self-efficacy items. However, it is also interesting to note that test-related self-efficacy measures might be used to finesse some of the difficulties we have raised, in relation to the longitudinal studies of

functional self-efficacy noted earlier. In well-designed longitudinal studies of achievement, standardized tests are designed both to assess achievement with items that are age-appropriate, and to provide a standardized measure of achievement along a common metric by using overlapping sets of items. To the extent that the self-efficacy items mirror the test items used on different occasions, self-efficacy responses might also be constructed so as to be age-appropriate and still provide measures along a common metric.

Conclusions: Directions for Future Research Based on Lessons Learned From Revisiting Marsh, Roche, Pajares, and Miller (1997)

We conclude by revisiting the calls for further research nominated by [Marsh et al. \(1997\)](#) in their dialogue/debate about distinctions between self-efficacy and self-concept, the role of evaluation and generalized measures of self-efficacy, and the appropriate construction of self-efficacy items. As noted in our earlier discussion, these self-efficacy and self-concept researchers called for research that:

- more fully delineates apparently overlapping constructs such as self-efficacy and self-concept on grounds other than domain specificity. This was, perhaps, the overarching focus of the present investigation, highlighting the relevance of jingle-jangle fallacies and our a priori hypotheses distinguishing between academic self-concept and self-efficacy.
- explores the evaluative component of self-efficacy responses, which seems to be important to the ability of self-efficacy beliefs to guide future behavior, but which is attenuated in operationalizations of self-efficacy in much educational research. Here we took a somewhat different perspective, arguing that, consistently with classic self-efficacy theory, appropriately constructed self-efficacy items should not have an evaluative component (also see earlier discussion in relation to [Williams & Rhodes, 2016](#)). Indeed, for us, this is an important distinction between self-efficacy and self-concept.
- evaluates further the theoretical and practical implications of more generalized or global self-efficacy measures, which seem antithetical to self-efficacy theory as originally conceived. Here we took a more nuanced perspective, distinguishing between what we referred to as generalized self-efficacy measures (which we argued to be more self-concept-like measures that should not be labeled as self-efficacy) and potentially more appropriate global measures, based on multiple task-specific self-efficacy items. However, we also identified a number of challenges to the construction of self-efficacy measures that attempt to bridge this gap between highly task-specific measures that were the historical basis of self-efficacy research, and more domain-general measures.
- is based on longitudinal data. Here, our focus was on the use of longitudinal data to better distinguish between different self-belief constructs. With regard to pretest variables, we showed that self-concept and self-efficacy differed in terms of frame-of-reference (social and dimensional comparison) effects. In a multitrait-multimethod analysis of test-retest data for the different self-belief constructs, we provided support

for our a priori classification of constructs as self-concept-like and self-efficacy-like factors. With regard to posttest variables, we showed that after controlling for pretest variables, self-concept responses were consistently more strongly related to important outcomes (achievement, aspirations) than were self-efficacy responses.

Our focus on frame-of-reference effects was not the major focus of Marsh et al. (1997). However, that paper did emphasize that self-efficacy and self-concept responses were likely to differ in relation to frame-of-reference effects and that this was related to the role of evaluation, which is central to self-concept responses but attenuated in self-efficacy responses. Expanding upon these suggestions, here we provide empirical support for this important distinction between self-concept and self-efficacy responses. Indeed, ours is apparently the first study to propose in a single theoretical model, and to demonstrate within a single statistical model, that there are consistently negative effects of dimensional (I/E) and social (BFLPE) comparisons for self-concept responses that are completely attenuated for self-efficacy responses.

Generalizability of Conclusions: Limitations and Directions for Further Research

In the last section, we posed potentially far-reaching conclusions about the nature of the murky distinction between self-concept and self-efficacy that provide a heuristic framework for further research. Nevertheless, it is also important to evaluate the generalizability of our conclusions in relation to limitations in the present investigation that provide direction for further research.

A potential limitation and direction for further research is to test the generalizability of our conclusions, which are based on responses by German secondary school students, to other countries, cultures, educational systems, and levels of education. A significant direction for further research is to evaluate the generalizability of our conclusions about the murky distinctions between academic self-concept, generalized self-efficacy, outcome expectancies, test-related self-efficacy, and functional self-efficacy, with other instruments designed to measure these and related self-belief constructs. Thus, for example, we have considered only one of the many scales designed to measure expectancy of success, although we anticipate that our results will generalize to other measures of academic self-concept and expectancy of success constructs that have a similar level of domain specificity (i.e., to math in the present investigation). However, we also note that in educational psychology in particular there is a plethora of seemingly related constructs that arise from apparently different theoretical perspectives and are given distinctively different labels. This lack of clarity in the key constructs underpinning much educational psychology research undermines conceptual understanding of the critical features of the different measures and the ability to synthesize research across the different measures. The conceptual and methodological focus of the present investigation should provide a useful starting point for such research, to clarify not only the murky distinction between self-concept and self-efficacy, but also, potentially, a wide range of other constructs in educational psychology and psychology more generally.

References

- Atkinson, J. W. (1964). *An introduction to motivation*. Oxford, England: Van Nostrand.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215. <http://dx.doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*, 1175–1184. <http://dx.doi.org/10.1037/0003-066X.44.9.1175>
- Bandura, A. (1994). Social cognitive theory and exercise of control over HIV infection. In R. J. DiClemente & J. L. Peterson (Eds.), *AIDS prevention and mental health. Preventing AIDS: Theories and methods of behavioral interventions* (pp. 25–59). Reprinted in modified form from R. DiClemente (Ed.), *Adolescents and AIDS: A Generation in Jeopardy*, Newbury Park, CA: Sage, 1992, pp. 89–116). http://dx.doi.org/10.1007/978-1-4899-1193-3_3
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman/Times Books/Henry Holt & Co.
- Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based majors. *Journal of Vocational Behavior*, *23*, 329–345. [http://dx.doi.org/10.1016/0001-8791\(83\)90046-5](http://dx.doi.org/10.1016/0001-8791(83)90046-5)
- Betz, N. E., & Hackett, G. (1993). *Manual for the mathematics self-efficacy scale*. Redwood City, CA: Mind Garden.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187–215. <http://dx.doi.org/10.1037/0033-2909.117.2.187>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley. <http://dx.doi.org/10.1002/9781118619179>
- Bong, M. (1996). Problems in academic motivation research and advantages of their solutions. *Contemporary Educational Psychology*, *21*, 149–165. <http://dx.doi.org/10.1006/ceps.1996.0013>
- Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, *34*, 139–153. http://dx.doi.org/10.1207/s15326985ep3403_1
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*, 1–40. <http://dx.doi.org/10.1023/A:1021302408382>
- Brunswick, E. (1952). *The conceptual framework of psychology*. Oxford, England: University Chicago Press.
- Byrne, B. M. (1996). Measuring self-concept: The conceptual issues. In B. M. Byrne (Ed.), *Measurement and instrumentation in psychology. Measuring self-concept across the life span: Issues and instrumentation* (pp. 1–35).
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. <http://dx.doi.org/10.1037/h0046016>
- Campbell, D. T., & O'Connell, E. J. (1967). Methods factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, *2*, 409–426. http://dx.doi.org/10.1207/s15327906mbr0204_1
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. <http://dx.doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5
- Eastman, C., & Marzillier, J. S. (1984). Theoretical and methodological difficulties in Bandura's self-efficacy theory. *Cognitive Therapy and Research*, *8*, 213–229. <http://dx.doi.org/10.1007/BF01172994>
- Eccles, J. (1984). Sex differences in achievement patterns. *Nebraska Symposium on Motivation*, *32*, 97–132.

- Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, *11*, 135–172. <http://dx.doi.org/10.1111/j.1471-6402.1987.tb00781.x>
- Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, *44*, 78–89. <http://dx.doi.org/10.1080/00461520902832368>
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*, 215–225. <http://dx.doi.org/10.1177/0146167295213003>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132. <http://dx.doi.org/10.1146/annurev.psych.53.100901.135153>
- Elliot, A. J., & Dweck, C. S. (2005). Competence and motivation: Competence as the core of achievement motivation. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 3–12). New York, NY: Guilford Press Publications.
- Enders, C. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in 3 western European societies: England, France and Sweden. *The British Journal of Sociology*, *30*, 415–441. <http://dx.doi.org/10.2307/589632>
- Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*, 519–533. <http://dx.doi.org/10.1037/0012-1649.38.4.519>
- Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, *101*, 705–716. <http://dx.doi.org/10.1037/a0014695>
- Frenzel, A. C., Pekrun, R., Dicke, A. L., & Goetz, T. (2012). Beyond quantitative decline: Conceptual shifts in adolescents' development of interest in mathematics. *Developmental Psychology*, *48*, 1069–1082. <http://dx.doi.org/10.1037/a0026895>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. <http://dx.doi.org/10.1146/annurev.psych.58.110405.085530>
- Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2015). Directionality of the associations of high school expectancy-value, aspirations, and attainment: A longitudinal study. *American Educational Research Journal*, *52*, 371–402. <http://dx.doi.org/10.3102/0002831214565786>
- Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction*, *49*, 81–91. <http://dx.doi.org/10.1016/j.learninstruc.2016.12.007>
- Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*, 1163–1176. <http://dx.doi.org/10.1037/a0039440>
- Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition*, *4*, 119–151. <http://dx.doi.org/10.1521/soco.1986.4.2.119>
- Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 553–617). Hoboken, NJ: Wiley.
- Harter, S. (2012). *The construction of the self: Developmental and socio-cultural foundations*. New York, NY: Guilford Press.
- Hattie, J. (1992). *Self-concept*. Hillsdale, NJ: Erlbaum.
- Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. *Motivation and Emotion*, *16*, 231–247. <http://dx.doi.org/10.1007/BF00991653>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*, 505–528. <http://dx.doi.org/10.1016/j.jsp.2011.07.001>
- Huang, C. (2012). Discriminant and incremental validity of self-concept and academic self-efficacy: A meta-analysis. *Educational Psychology*, *32*, 777–805. <http://dx.doi.org/10.1080/01443410.2012.732386>
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527. <http://dx.doi.org/10.1111/1467-8624.00421>
- James, W. (1963). *The principles of psychology* (Vol. 2). New York, NY: Holt. (Original work published 1890)
- Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*, 1195–1199. <http://dx.doi.org/10.1037/a0015665>
- Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal investigations. In J. R. Nesselrode & B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–351). New York, NY: Academic Press.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. Oxford, England: World Book Co.
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, *19*, 355–365. <http://dx.doi.org/10.1016/j.lindif.2008.10.009>
- Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. (1944). Level of aspiration. In J. M. Hunt (Ed.), *Personality and the behavior disorders* (pp. 333–378). Oxford, England: Ronald Press.
- Maddux, J. E. (2009). Self-efficacy: The power of believing you can. In S. J. Lopez & C. R. Snyder (Eds.), *Oxford library of psychology: Oxford handbook of positive psychology* (pp. 335–343). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780195187243.013.0031>
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*, 129–149. <http://dx.doi.org/10.3102/00028312023001129>
- Marsh, H. W. (1988). Multitrait multimethod analysis. In J. P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: An international handbook* (pp. 570–580). Oxford, England: Pergamon Press.
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*, 623–636. <http://dx.doi.org/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. <http://dx.doi.org/10.3102/00028312028002445>
- Marsh, H. W. (1993). Academic self-concept: Theory, measurement, and research. In J. M. Suls (Ed.), *The self in social perspective* (pp. 59–98). Hillsdale, NJ: Erlbaum.
- Marsh, H. W. (1994). Sport motivation orientations: Beware of the jingle-jangle fallacies. *Journal of Sport & Exercise Psychology*, *16*, 365–380. <http://dx.doi.org/10.1123/jsep.16.4.365>

- Marsh, H. W. (1995). The analysis of multitrait multimethod data. In T. H. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.). Oxford, UK: Pergamon Press.
- Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. London, England: British Psychological Society.
- Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of U.S. and Saudi Arabian TIMSS responses. *Journal of Cross-Cultural Psychology, 45*, 777–804. <http://dx.doi.org/10.1177/0022022113519858>
- Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. A. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology, 107*, 258–271. <http://dx.doi.org/10.1037/a0037485>
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163. <http://dx.doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., Craven, R. G., Hinkley, J. W., & Debus, R. L. (2003). Evaluation of the Big-Two-Factor Theory of academic motivation orientations: An evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research, 38*, 189–224. http://dx.doi.org/10.1207/S15327906MBR3802_3
- Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the Self-Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment, 17*, 81–102. <http://dx.doi.org/10.1037/1040-3590.17.1.81>
- Marsh, H. W., & Grayson, D. (1995). Latent-variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues and applications* (pp. 177–198). Thousand Oaks, CA: Sage.
- 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). Oxford, England: Wiley.
- Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education, 64*, 364–390. <http://dx.doi.org/10.1080/00220973.1996.10806604>
- Marsh, H. W., & Hau, K.-T. (1998). Is parsimony always desirable: Response to Hoyle, Sivo & Willson, Markus, Mulaik, Tweedledee, Tweedledum, the Cheshire Cat, and others. *Journal of Experimental Education, 66*, 274–285. <http://dx.doi.org/10.1080/00220979809604412>
- 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. <http://dx.doi.org/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. <http://dx.doi.org/10.1037/0022-0663.96.1.56>
- Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology, 32*, 151–170. <http://dx.doi.org/10.1016/j.cedpsych.2006.10.008>
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Erlbaum.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999). Findings. *Structural Equation Modeling, 11*, 320–341. http://dx.doi.org/10.1207/s15328007sem1103_2
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*, 107–117. <http://dx.doi.org/10.1037/0021-9010.73.1.107>
- 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. <http://dx.doi.org/10.3102/00028312038002321>
- Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction, 33*, 50–66. <http://dx.doi.org/10.1016/j.learninstruc.2014.04.002>
- Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology, 39*, 326–341. <http://dx.doi.org/10.1016/j.cedpsych.2014.08.003>
- Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods, 18*, 257–284. <http://dx.doi.org/10.1037/a0032773>
- Marsh, H. W., Martin, A. J., & Jackson, S. (2010). Introducing a short version of the physical self-description questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport & Exercise Psychology, 32*, 438–482. <http://dx.doi.org/10.1123/jsep.32.4.438>
- Marsh, H. W., Martin, A. J., Yeung, A. S., & Craven, R. C. (2017). Competence self-perceptions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: theory and application* (2nd ed., pp. 85–115). New York, NY: Guilford Press.
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439–476. <http://dx.doi.org/10.1080/10705510903008220>
- 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. <http://dx.doi.org/10.1177/0146167207312313>
- Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*, 213–231. <http://dx.doi.org/10.1037/0022-3514.47.1.213>
- Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology, 54*, 263–280. <http://dx.doi.org/10.1037/dev0000393>
- Marsh, H. W., Pekrun, R., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109*, 425–438. <http://dx.doi.org/10.1037/edu0000144>
- Marsh, H. W., Roche, L. A., Pajares, F., & Miller, D. (1997). Item-specific efficacy judgments in mathematical problem solving: The downside of standing too close to trees in a forest. *Contemporary Educational Psychology, 22*, 363–377. <http://dx.doi.org/10.1006/ceps.1997.0942>

- Marsh, H. W., & Seaton, M. (2015). The big-fish-little-pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. In A. J. Elliott (Ed.), *Advances in motivation science* (Vol. 2, pp. 127–184). New York, NY: Elsevier. <http://dx.doi.org/10.1016/bs.adms.2015.05.002>
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, 20, 319–350. <http://dx.doi.org/10.1007/s10648-008-9075-6>
- Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20, 107–123. http://dx.doi.org/10.1207/s15326985ep2003_1
- Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, 44, 631–669. <http://dx.doi.org/10.3102/0002831207306728>
- Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, 100, 510–524. <http://dx.doi.org/10.1037/0022-0663.100.3.510>
- Marsh, H. W., Walker, R., & Debus, R. (1991). Subject-specific components of academic self-concept and self-efficacy. *Contemporary Educational Psychology*, 16, 331–345. [http://dx.doi.org/10.1016/0361-476X\(91\)90013-B](http://dx.doi.org/10.1016/0361-476X(91)90013-B)
- Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (pp. 427–458). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/13273-015>
- Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, 82, 60–70. <http://dx.doi.org/10.1037/0022-0663.82.1.60>
- Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120, 544–560. <http://dx.doi.org/10.1037/a0032459>
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79, 1129–1167. <http://dx.doi.org/10.3102/0034654309337522>
- Murayama, K., Pekrun, R., Lichtenfeld, S., & vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development*, 84, 1475–1490. <http://dx.doi.org/10.1111/cdev.12036>
- Murayama, K., Pekrun, R., Suzuki, M., Marsh, H. W., & Lichtenfeld, S. (2016). Don't aim too high for your kids: Parental overaspiration undermines students' learning in mathematics. *Journal of Personality and Social Psychology*, 111, 766–779. <http://dx.doi.org/10.1037/pspp0000079>
- Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology*, 25, 3–53. <http://dx.doi.org/10.1006/ceps.1999.1019>
- Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, 60, 489–503. <http://dx.doi.org/10.1007/BF02294325>
- Muthén, L. K., & Muthén, B. (2008–2014). *Mplus user's guide*. Los Angeles, CA: Author.
- Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: The Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology*, 31, 629–656. <http://dx.doi.org/10.1080/01443410.2011.586416>
- Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, 104, 1033–1053. <http://dx.doi.org/10.1037/a0027697>
- Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, 22, 1058–1066. <http://dx.doi.org/10.1177/0956797611415540>
- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17, 372–411. <http://dx.doi.org/10.1177/1094428114548590>
- OECD. (2004). *PISA 2003 Tech. Rep. No.* Paris, France: Author.
- Pajares, F. (1996). Self-efficacy beliefs and mathematics problem-solving of gifted students. *Contemporary Educational Psychology*, 21, 325–344. <http://dx.doi.org/10.1006/ceps.1996.0025>
- Pajares, F. (2009). Toward a positive psychology of academic motivation: The role of self-efficacy beliefs. In R. Gilman, E. S. Huebner, & M. J. Furlong (Eds.), *Handbook of positive psychology in schools* (pp. 149–160). New York, NY: Routledge/Taylor & Francis Group.
- Pajares, F., & Miller, D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology*, 42, 190–198. <http://dx.doi.org/10.1037/0022-0167.42.2.190>
- Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, 34, 29–48. <http://dx.doi.org/10.1080/01443410.2013.797339>
- Parker, P. D., Marsh, H. W., Guo, J., Anders, J., Shure, N., & Dicke, T. (2018). An information distortion model of social class differences in math self-concept, intrinsic value, and utility value. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000215>
- Parker, P. D., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2013). Differential school contextual effects for math and English: Integrating the big-fish-little-pond effect and the internal/external frame of reference. *Learning and Instruction*, 23, 78–89. <http://dx.doi.org/10.1016/j.learninstruc.2012.07.001>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315–341. <http://dx.doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, 88, 1653–1670. <http://dx.doi.org/10.1111/cdev.12704>
- Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). New York, NY: Taylor & Francis.
- Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 17–37). Münster, Germany: Waxmann.
- Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (pp. 103–122). Hoboken, NJ: Wiley.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivated Strategies for Learning Questionnaire. *Journal of Educational Psychology*, 82, 33–40.

- Schunk, D. H., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). New York, NY: Guilford Press Publications.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor, United Kingdom: NFER-NELSON.
- Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond-effect: Generalizability and moderation—Two sides of the same coin. *American Educational Research Journal*, *47*, 390–433. <http://dx.doi.org/10.3102/0002831209350493>
- Seifert, T. (2004). Understanding student motivation. *Educational Research*, *46*, 137–149. <http://dx.doi.org/10.1080/0013188042000222421>
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*, 407–441. <http://dx.doi.org/10.3102/00346543046003407>
- Skaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. *Journal of Educational Psychology*, *82*, 546–554. <http://dx.doi.org/10.1037/0022-0663.82.3.546>
- Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist*, *37*, 233–244. http://dx.doi.org/10.1207/S15326985EP3704_3
- Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology*, *71*, 549–570. <http://dx.doi.org/10.1037/0022-3514.71.3.549>
- Tolman, E. C. (1932). *Purposive behavior in animals and men*. Oxford, England: Appleton-Century.
- Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy-value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*, 763–777. <http://dx.doi.org/10.1037/a0027470>
- 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. <http://dx.doi.org/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*, 37–41. http://dx.doi.org/10.1207/s15326985ep3902_3
- vom Hofe, R., Kleine, M., Blum, W., & Pekrun, R. (2005). On the role of “Grundvorstellungen” for the development of mathematical literacy: First results of the longitudinal study PALMA. *Mediterranean Journal for Research in Mathematics Education*, *4*, 67–84.
- vom Hofe, R., Pekrun, R., Kleine, M., & Götz, T. (2002). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA): Konstruktion des Regensburger Mathematikleistungstests für 5.-10. Klassen [Project for the analysis of learning and achievement in mathematics (PALMA): Development of the Regensburg mathematics achievement test for grades 5 to 10]. *Zeitschrift für Pädagogik*, *45*, 83–100.
- Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*, 265–310. [http://dx.doi.org/10.1016/0273-2297\(92\)90011-P](http://dx.doi.org/10.1016/0273-2297(92)90011-P)
- Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Social, emotional, and personality development* (pp. 933–1002). Hoboken, NJ: Wiley.
- Wigfield, A., Tonks, S., & Klauda, S. L. (2016). Expectancy-value theory. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (pp. 55–74). New York, NY: Routledge.
- Williams, D. M., & Rhodes, R. E. (2016). The confounded self-efficacy construct: Conceptual analysis and recommendations for future research. *Health Psychology Review*, *10*, 113–128. <http://dx.doi.org/10.1080/17437199.2014.941998>
- Wylie, R. C. (1989). *Measures of self-concept*. Lincoln, NE: University of Nebraska Press.
- Zell, E., & Alicke, M. D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. *Journal of Personality and Social Psychology*, *97*, 467–482. <http://dx.doi.org/10.1037/a0015453>
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*, 82–91. <http://dx.doi.org/10.1006/ceps.1999.1016>

Received September 21, 2017

Revision received February 22, 2018

Accepted March 3, 2018 ■