Contemporary Educational Psychology 41 (2015) 13–24

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes

Malte Jansen [a,*,1], Ronny Scherer [b], Ulrich Schroeders [c]

[a] *Institute for Educational Quality Improvement (IQB), Humboldt University of Berlin, Berlin, Germany*
[b] *Centre for Educational Measurement, University of Oslo (CEMO), Oslo, Norway*
[c] *Department of Educational Science, University of Bamberg, Bamberg, Germany*

ARTICLE INFO

Article history:
Available online 11 November 2014

Keywords:
Academic self-concept
Self-efficacy
Science education
Big-Fish-Little-Pond-Effect
Future-oriented motivation
Inquiry-based teaching

ABSTRACT

Self-concept and self-efficacy are two of the most important motivational predictors of educational outcomes. As most research has studied these constructs separately, little is known about their differential relations to peer ability, opportunities-to-learn in classrooms, and educational outcomes. We investigated these relations by applying (multilevel) structural equation modeling to the German PISA 2006 data set. We found a correlation of $\rho = .57$ between self-concept and self-efficacy in science, advocating distinguishable constructs. Furthermore, science self-concept was better predicted by the average peer achievement (*Big-Fish-Little-Pond Effect*), whereas science self-efficacy was more strongly affected by inquiry-based learning opportunities. There were also differences in the predictive potential for educational outcomes: Self-concept was a better predictor of future-oriented motivation to aspire a career in the sciences, whereas self-efficacy was a better predictor of current ability. The study at hand provides strong evidence for the related but distinct nature of the two constructs and extends existing research on students' competence beliefs toward social comparisons and opportunities-to-learn. Further implications for the relevance of inquiry-based classroom activities and for the assessment of competence beliefs are discussed.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Students' competence beliefs, that is, their academic self-concept (ASC) and self-efficacy (ASE), are positively related to various desirable scholastic outcomes such as achievement, effort, and attainment (Huang, 2011; Marsh & Martin, 2011; Nagengast et al., 2011; Pajares, 1996; Swann, Chang-Schneider, & Larsen McClarty, 2007; Trautwein, Lüdtke, Schnyder, & Niggli, 2006; Valentine, DuBois, & Cooper, 2004). Similarities but also meaningful differences in the conceptualizations and operationalizations of the two constructs have been pointed out (Bong & Clark, 1999; Bong & Skaalvik, 2003). However, studies have rarely investigated whether the two constructs show differential relations to antecedents of competence beliefs and educational outcomes or are merely different labels for the same construct (*jangle-fallacy*, Kelley, 1927; Marsh, Craven, Hinkley, & Debus, 2003; Reschly & Christenson, 2012). More precisely, little is known about the potentially differential effects of social comparisons with peers and instructional activities such as inquiry-based learning on competence beliefs. Further, only few studies have focused on these differences in the domain of science even though science education in high-school classrooms may offer different antecedents of competence beliefs when compared with other subjects. Moreover, predicting educational outcomes in the sciences using motivational factors is seen as critically important in both research and educational policy (Bybee & McCrae, 2011, OECD, 2007; Osborne, Simon, & Collins, 2003; Taskinen, Schütte, & Prenzel, 2013; Tsai, Jessie Ho, Liang, & Lin, 2011; Wigfield & Eccles, 2000).

The main goals of the present study were to analyze the empirical differences between ASC and ASE in science and to disentangle their potential differential relations to a selected set of covariates and outcomes. In the following, we will first describe the conceptual differences between ASC and ASE that have been identified in theoretical reviews. Subsequently, we will sketch a conceptual research model for studying this distinction that describes the assumed relations between social comparisons and opportunities-to-learn as sources of competence beliefs, competence beliefs, and educational outcomes. From this model and previous empirical research on the relation between ASE and ASC, we derived and tested specific hypotheses.

1.1. Conceptual differences between self-concept and self-efficacy

In their reviews, Bong and colleagues described similarities and differences in the conceptualizations and operationalizations of ASE and ASC in detail (Bong, Cho, Ahn, & Kim, 2012; Bong & Clark, 1999;

* Corresponding author. German Institute for International Educational Research (DIPF), Warschauer Str. 34–38, 10243 Berlin, Germany.
E-mail address: jansen@dipf.de (M. Jansen).
[1] Malte Jansen is now at the German Institute for International Educational Research (DIPF), Berlin.

http://dx.doi.org/10.1016/j.cedpsych.2014.11.002
0361-476X/© 2014 Elsevier Inc. All rights reserved.

Bong & Skaalvik, 2003). Among others, the authors identified differences with regard to the nature and the sources of judgment on which the competence beliefs were based and their predictive validity for certain outcomes. These differences will be briefly summarized in the following paragraphs.

Broadly speaking, ASC refers to the self-evaluation of one's general ability in a domain (Marsh & Martin, 2011). ASC is assumed to be a relatively stable, multidimensional, hierarchical, and domain-specific construct and is most commonly studied at the level of school subjects such as mathematics, English, and science (Marsh, 1990; Shavelson, Hubner, & Stanton, 1976). Even though based on "objective" achievement feedback such as grades, ASC is still a subjective evaluation of one's own achievement (Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014). For example, when students respond to an item such as "I do well in science", their interpretation of what "well" means will vary based on their own standards and frames of references. Therefore, different comparison processes using different frames of reference are assumed to the most important sources of ASCs (Möller & Marsh, 2013). At least three distinct comparison processes affect self-concept development: First, students compare their performance in one domain with the performance of their peers in the same domain (*social comparisons*; Festinger, 1954; Marsh, 1987; Seaton, Marsh, & Craven, 2009). Second, students compare their performance in one domain with their previous performance in the same domain (*temporal comparisons*; Albert, 1977; Möller, 2005). Third, they compare their performance in one domain with their own performance in other domains thus developing a profile of self-perceived strengths and weaknesses (*dimensional comparisons*; Marsh, 1986; Marsh et al., 2015; Möller & Marsh, 2013). Even though considerations of comparison processes have guided much of self-concept research, it should be noted that there are further influential sources of ASCs such as teachers' and parents' appraisal or stereotype endorsement with respect to gender and ethnicity (Kessels & Hannover, 2008; Okeke, Howard, Kurtz-Costes, & Rowley, 2009; Spinath & Spinath, 2005; Tiedemann, 2000).

ASE, on the other hand, refers to a student's perception of his or her ability to successfully complete a specific academic task or reach an academic goal (Pajares, 1996). According to Bandura (1997), self-efficacy refers to 'beliefs in one's capabilities to organize and execute the courses of action required to manage prospective situations' (p. 2). It is considered a multidimensional task- and domain-specific construct, but somewhat less stable and hierarchical compared with ASC (Bong & Skaalvik, 2003). Furthermore, it is assumed to influence the choice and pursuance of tasks and actions. Compared with ASC, ASE more strongly relies on specific judgments of whether a task can be carried out successfully or a goal can be achieved. There are four major sources of self-efficacy (Bandura, 1997; Usher & Pajares, 2009). If students are asked to estimate their probability of successfully solving a math problem, their past experience with similar math problems will be the most informative source of judgment (Usher & Pajares, 2009). Conceptually speaking, students' ASE is most strongly affected by *mastery experiences*, whereas social and dimensional comparisons play subordinate roles (Pajares, 1996; Usher & Pajares, 2009). These experiences are considered the most powerful source of self-efficacy and occur when students successfully complete academic tasks and achieve goals. However, students' social context is also assumed to affect the development of ASE through two path ways, especially when students do not have sufficient experience with a given task to draw on previous experience (Usher & Pajares, 2009). *Vicarious experiences* result from the observation of peers performing certain tasks; students use this as a source of information for their own ability to perform that specific task ("If someone like me can do it, I might be able to do it as well"). *Social persuasions* refer to encouragement from peers, teachers, and parents which can enhance students' confidence in performing certain tasks. Finally, their own *physiological state* (i.e., perceived stress or anxiety when confronted with a task) is assumed to be an additional source of self-efficacy as students "learn to interpret their physiological arousal as an indicator of personal competence by evaluating their own performances under differing conditions" (Usher & Pajares, 2009; p. 90).

Both competence beliefs are assumed not only to be affected by ability (*skill development model*; Helmke & van Aken, 1995) but also to affect ability (*self-enhancement model*) and academic achievement (e.g., school grades, standardized test scores, grade retention). Due to the conceptual distinction between ASC and ASE, differences in their relations to educational outcomes have been assumed (Bong & Skaalvik, 2003). Besides affecting achievement (see "reciprocal effects model"; Marsh & Martin, 2011), ASC should also influence course choices, educational aspirations, and affective reactions such as school anxiety (Nagengast & Marsh, 2012). By contrast, ASE is assumed to affect motivational constructs such as goal orientations, goal setting, persistence, and task choices during learning processes in addition to educational achievement (e.g., Pajares, 1996; Pajares, Britner, & Valiante, 2000; Parker et al., 2014).

1.2. A research model for studying the empirical distinction between self-concept and self-efficacy

The conceptual differences described so far not only imply that ASC and ASE are distinct constructs both in their conceptualization and measurement, but may also rely on different sources of information and show differential relations to educational outcomes. A research model for studying the distinction between ASC and ASE should therefore account for these aspects. In our research model (see Fig. 1), we first assume that ASC and ASE can be separated in measurement. On the basis of this assumption, we focus on selected antecedents that have rarely been taken into account when juxtaposing the two competence beliefs. More precisely, we study the effects of social comparisons and the opportunities-to-learn in science classrooms on ASC and ASE. These characteristics of the learning environment include classroom activities that might evoke mastery experiences and the average ability of peers. Peers serve both as a social frame of references for comparing one's own ability and as a source of vicarious experiences. Third, we juxtapose ASC and ASE according to their predictive potential toward educational outcomes such as science achievement and students' future-oriented motivation in science. These outcomes can be regarded as cognitive and motivational indicators of school success. In the following, we describe the research model in more detail by presenting empirical evidence for the relations assumed in the model.

1.2.1. Relation between ASC and ASE

The evidence with regard to the factor structures of ASC and ASE is inconclusive, including the separability of the two constructs as well as the magnitude of their relation. Some studies found high correlations indicating that a distinct assessment of the two constructs might be difficult (Bong et al., 2012; Marsh, Dowson, Pietsch, & Walker, 2004), whereas others found only moderate correlations (Ferla, Valcke, & Cai, 2009; Ferla, Valcke, & Schuyten, 2010). As evident from these large disparities, the operationalization of the constructs is crucial and affects their degree of relatedness. The operationalization of ASC is relatively straightforward and agreed upon. The questionnaire items used in large-scale assessment studies such as PISA or TIMSS rely on established items identically or similarly worded as well-established questionnaires such as the Self-Description-Questionnaire. For example, items that represent general self-evaluations of ability in a domain such as "*I learn quickly in [domain]*" (SDQ), "*I get good marks in [domain]*", "*I learn [domain] topics quickly*" (PISA 2006) or "*I usually do well in [domain]*" (TIMSS 2011).

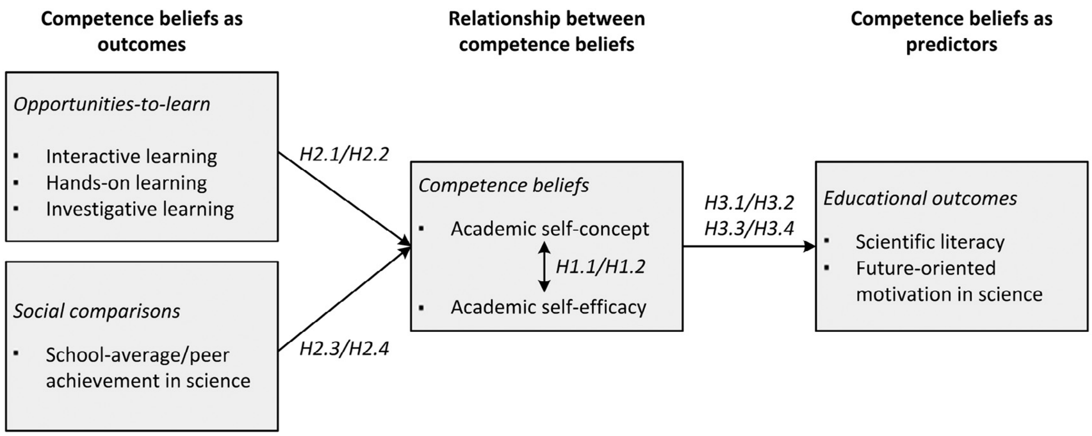

Fig. 1. Conceptual model guiding the juxtaposition of ASC and ASE in the present study.

However, operationalizations of ASE at the domain level are more controversially discussed. There are two major approaches that refer to Bandura's (2006) recommendations on how to construct scales to measure ASE. Although both approaches refer to the task-specific and goal-oriented nature of ASE, their definition of what constitutes a relevant "task" on domain level varies. In particular, the first approach relies on measuring domain-specific self-efficacy beliefs with items that refer to more general tasks or goals such as "doing well" in a class (Pintrich & De Groot, 1990) or getting a specific grade in a class (Zimmerman & Bandura, 1994). This kind of operationalization has been criticized, however, for not taking into account the specificity of different tasks within a domain. As Pajares (1996) put it, "domain-specific assessments, such as asking students to report their confidence to learn mathematics or writing [...] are inferior to task-specific judgments because the subdomains differ markedly in the skills required" (p. 547). Moreover, Skaalvik and Skaalvik (2004) argued that the differentiation between self-efficacy measured with this approach and domain-specific self-concepts is blurry.

In the second approach that is used in large-scale studies such as PISA and TIMSS, domain-specific self-efficacy beliefs are measured by items that refer to specific skills and contexts. These skills are based on the concepts of literacy in a specific domain. Because several different tasks are included, they are assumed to cover the depth and breadth of the skills needed to master the subject.

Bong et al. (2012) studied the relation between ASC and ASE by using two different self-efficacy measures. They showed that ASC and ASE in math, as measured by the *Motivated Strategies for Learning Questionnaire* (e.g., "*I expect to do well in this class*"; Pintrich & De Groot, 1990), could hardly be distinguished ($r = .91$), whereas the correlation was lower ($r = .78$) when an operationalization based on specific grades was used (e.g. "*I can get more than '80' in the math exam*"). Both operationalizations followed the first approach and did not refer to specific tasks. In other studies, where task-specific ASE items were implemented, the correlation with ASC was only moderate in the domain of math ($.37 < r < .41$; Ferla et al., 2009, 2010). In sum, the approaches of measuring ASE differ in their specificity of task descriptions which may lead to differential findings on the correlations between ASE and ASC.

1.2.2. Antecedents of competence beliefs

In addition to examining measurement models, comparative research on ASC and ASE aiming to back up their distinctiveness has also looked at the relations of the two constructs to antecedents of competence beliefs. These included student characteristics (e.g., gender, socio-economic status, prior achievement), as well as characteristics of the learning environment (Ferla et al., 2009; Marsh, 1987; Parker et al., 2014; Usher & Pajares, 2009). In our study, we focus on two antecedents of competence beliefs: peer ability as a proxy for social comparisons and classroom activities as proxies for opportunities-to-learn that may lead to mastery experiences in science.

Research showed that individual competence beliefs are influenced by social comparison processes in the classroom. It is assumed that students compare their abilities with the perceived ability of their average classmates (the "*generalized other*"; Huguet et al., 2009; Marsh, Kuyper, Morin, Parker, & Seaton, 2014). An often replicated finding of studies on social comparison is the "Big-Fish-Little-Pond-Effect" (BFLPE; Marsh, 1987), which describes the negative contextual effect of class- or school-aggregated ability on students' ASC after controlling for individual ability. This means that a student in a high ability class (a "big pond") would show a lower self-concept of his or her ability than an equally able student in a low ability class (a "big fish in a little pond"). This effect has been found in a variety of domains, grade levels, learning environments, and cultures and is regarded as evidence of the importance of social comparisons for self-concept development (e.g., Nagengast & Marsh, 2012; Seaton et al., 2009). Only one study examined the BFLPE on ASE and found a small negative effect (–.08 for general ASE compared with –.20 for general ASC; Marsh, Trautwein, Lüdtke, & Köller, 2008).

The classroom is not only a social environment, but also offers opportunities-to-learn and to acquire mastery experiences. Compared with individual learner characteristics, only few studies have dealt with the effects of classroom activities on competence beliefs even though such effects are highly plausible (Beghetto, 2007). Mastery experiences refer to students' previous success or failure in academic tasks. As Usher and Pajares (2009) pointed out, "Mastery experiences prove particularly powerful when individuals overcome obstacles or succeed on challenging tasks" (p. 89). Therefore, students would be more likely to experience mastery or failure if they are actively engaged in challenging tasks. Interactive and inquiry-based learning environments may offer more such opportunities than teacher-centered and less interactive environments because students are expected to participate actively in discussions, to engage in collaborative learning, and, in the sciences, carry out practical experiments (Freedman, 1997; Hofstein & Lunetta, 2004; Tsai et al., 2011). Therefore, if students succeed in completing inquiry-related tasks, these mastery experiences may lead to a higher ASE. For example, if the teacher asks the students to share their opinions in an interactive way or if students work on hands-on tasks such as designing and carrying out experiments, and they accomplish these tasks, they may feel more competent and confident in carrying out these tasks in future. Such inquiry-based learning approaches play a special role in the sciences because of their experimental nature (Taskinen et al., 2013).

1.2.3. *Predictive validity of competence beliefs for academic outcomes*

In addition to examining possible differential relations to antecedents of competence beliefs, we also wanted to determine whether ASC and ASE show empirical differences with regard to their effects on desirable academic outcomes. Among the most important outcomes are students' achievement and their educational choices and aspirations. Meta-analyses (Huang, 2011; Valentine et al., 2004) and longitudinal studies (Marsh & Martin, 2011; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005) have shown that the relations between academic achievement and the two competence beliefs are substantial and reciprocal. More specifically, there is evidence for positive associations between competence beliefs and achievement in the sciences (Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011; Nagengast et al., 2011). The studies that have compared ASC and ASE as predictors of standardized test scores in math have shown slightly higher effects for ASE (Ferla et al., 2009, 2010; Pajares & Miller, 1994).

ASC has also been shown to be positively related to educational choices and career aspirations (Nagengast & Marsh, 2012; Nagengast et al., 2011; Taskinen et al., 2013). There is less empirical evidence for ASE, but it points into the same direction (Betz, 2000). Parker et al. (2014) compared the effects of ASC and ASE on university course choice in a longitudinal study, in which participants, who began the study at the age of 15, were followed for 8 years. The authors found that only math ASC could predict whether a student would take STEM courses.

1.3. *The specifics of competence beliefs in science*

Previous research on the juxtaposition of ASC and ASE has mostly focused on math and language learning. Studies on their relation in science are rare, even though the sciences have received a lot of attention in both educational policy and research lately (Bybee & McCrae, 2011; OECD, 2007). Nevertheless, there are at least two arguments why previous research on the relation between ASC and ASE in math and languages may not be generalizable to the sciences.

First, competence beliefs are domain-specific. An impressive literature documents the domain-specificity of both ASC and ASE showing that domain-specific components of competence beliefs are better predictors of educational outcomes in a given domain than general competence beliefs (Marsh, 1990; Marsh & Craven, 2006; Pajares, 1996; Valentine et al., 2004). ASC and ASE should therefore be measured in domain-specific settings (Bong & Skaalvik, 2003; Mason, Boscolo, Tornatora, & Ronconi, 2013; Skaalvik & Skaalvik, 2004). As Bong et al. (2012; p. 339) put it: *"Because students develop unique understandings of not only themselves but also academic domains […], differential relations may exist among their self-perceptions and their motivation and achievement across different subject domains."* Hence, the generalizability of study results on competence beliefs from one domain to another cannot be taken for granted.

Second, competence beliefs are not only affected by teaching practices but also by students' beliefs about the domain at hand (e.g., Chen & Pajares, 2010; Mason et al., 2013). Epistemological beliefs (i.e., beliefs about knowledge and its certainty, simplicity, sources, and justification; Hofer & Pintrich, 1997) are inherently domain-specific because domains differ in their definition of knowledge. Therefore, domain-specific knowledge structures may lead to different epistemological beliefs which would then differentially affect competence beliefs. Further, there are also differences in student-perceived domain characteristics as shown by Goetz and colleagues (2014): In general, subjects from the quantitative domain such as math and physics and subjects from the verbal domain (e.g., English) are perceived very differently by students. For example, quantitative subjects are assumed to be more difficult, show less relations to everyday life and up-to-date topics compared with languages. But there are also differences in student-perceived domain characteristics between math and physics with physics showing a higher variety of content, a higher number of graphical illustrations of the material presented, and more relations to up-to-date topics. Math, however, was perceived as more wearisome but showing a higher coherence of content; further, good achievements in math were valued more highly by the students. If students take into account such different beliefs about school subjects as well as their own strengths and weaknesses in forming their competence beliefs, different competence beliefs in different domains are expected.

2. The present study: research questions and hypotheses

The present study takes a closer look at the empirical relation between ASC and ASE in the domain of science. As previously pointed out, this relation has not yet been studied in science and results are not easily generalizable. Thus, a replication of the factor-analytic distinction as well as the differential relations to academic outcomes is needed. However, our study also goes beyond replicating previous findings in a new domain. We examine potential differential effects of social comparisons (i.e., BFLPEs) and classroom activities on ASE and ASC. Based on our research model (Fig. 1), we thus explore the relation between the two constructs by focusing on (a) the factor structure of science self-concept and self-efficacy; (b) the antecedents of competence beliefs; (c) the prediction of educational outcomes by competence beliefs.

2.1. *The factorial structure of science self-concept and self-efficacy*

In the first step, we compared a one-factor model of academic competence beliefs in science with a two-factor model distinguishing between ASC and ASE as first-order factors. These analyses aimed to explore whether the two constructs could be empirically separated. Our hypotheses were as follows:

Hypothesis 1.1. The two-factor model distinguishing between ASC and ASE in science will show a superior fit to the data compared with a one-factor model.

Hypothesis 1.2. The two factors will be positively related.

2.2. *Competence beliefs as outcomes*

In a second step, we examined differential effects of predictors of competence beliefs. These included peer ability as a proxy for social comparisons and inquiry-related classroom activities as proxies for opportunities-to-learn. Since students use social comparisons to form their self-evaluations, we expected a negative effect of the average peer ability on both competence beliefs. Because social frames of reference are considered to be more important for the development of ASC than for ASE, we expected the potential negative effect of the average class ability to be weaker for self-efficacy.

With regard to ASE, having high-ability peers may not only have a negative effect (as a result of social comparisons), but also have a positive effect (as a result of vicarious experiences) on ASE. These effects are not separable and may cancel each other out. In line with this assumption, extensive research on the BFLPE has conclusively shown the overall generalizability of a substantial negative effect of peer ability on ASC (Nagengast & Marsh, 2012; Seaton et al., 2009). By contrast, Marsh et al. (2008) found only a weak BFLPE for ASE.

Mastery experiences in solving scientific tasks and especially in scientific experimenting are mainly acquired in science lessons. Inquiry-based learning environments characterized by opportunities for scientific investigation and "hands-on" activities such as experiments should thus affect students' competence beliefs in science. Given the higher importance of mastery experiences for ASE development compared with ASC, these effects were expected to be stronger for ASE. Our hypotheses regarding the effects of the social context and opportunities to learn in the classroom were thus:

Hypothesis 2.1. Inquiry-based activities in science classrooms will be positively associated with students' competence beliefs.

Hypothesis 2.2. This positive effect will be stronger for ASE compared with ASC.

Hypothesis 2.3. After controlling for individual ability, average school ability will show a negative effect on individual ASC (the BFLPE).

Hypothesis 2.4. This negative BFLPE will be weaker or even nonexistent for ASE.

2.3. Competence beliefs as predictors

Given the plethora of evidence on positive relations between competence beliefs and academic achievement (e.g., Huang, 2011), we expected positive relations in our study. We further hypothesized that the effects of ASE would be higher, even though there is some conflicting evidence (e.g., Choi, 2005). We also expected a positive effect of competence beliefs on students' motivation to enroll in a science course at university or to pursue a science career. Because career choices require a global assessment of one's competence in a domain, social and dimensional comparisons are important sources of information for such assessments. Moreover, since there is strong empirical evidence for the effects of ASC on educational choices in general and in the sciences in particular (e.g., Taskinen et al., 2013), we expected a higher predictive effect of ASC on students' future-oriented motivation to aspire a career in the sciences. Our hypotheses with regard to differential effects of competence beliefs on educational outcomes were consequently:

Hypothesis 3.1. ASC and ASE will show positive effects on science achievement.

Hypothesis 3.2. The positive effect on achievement will be higher for ASE compared with ASC.

Hypothesis 3.3. ASC and ASE will show positive effects on future-oriented motivation.

Hypothesis 3.4. The positive effect on future-oriented motivation will be stronger for ASC compared with ASE.

3. Method

3.1. Sample and procedure

We used data from the nationally representative sample of the PISA 2006 large-scale assessment in Germany. This sample included $N = 4{,}891$ high-school students (49.7% female, 80.7% native Germans) from 226 schools. Students' mean age was 15.9 years ($SD = 0.3$). Most of the students attended ninth grade (55.2%), but there were also students in 10th (28.6%), 8th (11.9%), 7th (1.4%) and 11th (.3%) grade. In the German school system, students change from elementary school to a tracked secondary school system. In our study, 39.2% of the participating students attended the high academic track (*Gymnasium*) whereas 60.8% attended the lower track, middle track or mixed-track schools. In two test sessions of 120 min each, assessments of mathematical, scientific, and reading literacy and a student background questionnaire were administered.

3.2. Measures

In the student questionnaire, a variety of cognitive, personality, classroom-, and school-related constructs were assessed, including the covariates of competence beliefs that were selected for the investigation at hand. To evaluate the reliability of the latent factors, we used McDonald's ω (1999).

3.2.1. Academic self-concept in science (ASC)

Self-concept in science was assessed using six Likert-type items. Students were asked to indicate the degree to which they agreed with statements such as "*I learn school science topics quickly*" or "*I can easily understand new ideas in school science*" ranging from 0 (*I strongly disagree*) to 3 (*I strongly agree*). These items were used in various studies (e.g., Jansen, Schroeders, & Lüdtke, 2014; Nagengast & Marsh, 2012; Scherer, 2013; Taskinen et al., 2013) and are highly similar to the items used in the SDQ and the ASDQ, the most validated self-concept measures (Marsh, 1990). The reliability of the self-concept scale was excellent ($\omega = .90$).

3.2.2. Science self-efficacy (ASE)

Science self-efficacy was measured with eight items for which students had to indicate how confident they felt when performing specific scientific tasks. These tasks comprised selected real-world science tasks that were closely related to scientific literacy as conceptualized in the PISA science framework (e.g., *predict how changes in an environment will affect the survival of certain species*). Items had to be rated with respect to students' mastery on a 4-point scale ranging from 0 (*I couldn't do this*) to 3 (*I could do this easily*). The reliability of the ASE scale was $\omega = .82$. On the basis of Bong and Skaalvik's (2003) comprehensive review of competence beliefs, we argue that the items can be regarded as proxies of students' self-efficacy for a number of reasons. First, the items are designed in such a way that they refer to future-oriented instead of past-oriented beliefs in mastering tasks. Second, they contain specific tasks, which are composed of specific competences and contents that are aligned with the notion of scientific literacy. Hence, from a conceptual perspective, we conclude that these items reflect students' self-efficacy. Further, Nagengast and Marsh (2014) provided empirical evidence for their validity.

3.2.3. Opportunities-to-learn in science lessons

The opportunities-to-learn in science classrooms were assessed by using a 4-point scale by which students had to report the frequency of specific inquiry-based activities (0 = *never or hardly ever*, 3 = *in all lessons*). The construction rationale for the items was based on a conceptual model of teaching and learning in science proposed by Seidel and Prenzel (2006). In this model, it is assumed that active, student-centered learning activities emphasizing student–teacher interactions, collaborative learning, classroom discussions, and hands-on activities facilitate students' literacy and interest in science. The activities were described as detailed as possible so that students could report the frequency of their occurrence reliably. The 11 items referred to three different facets of learning activities: (a) interactive and student-centered teaching and learning (e.g., "*Students have discussion about the topics*"; four items, $\omega = .72$), (b) hands-on activities in science (e.g., "*Students spend time in the laboratory to do practical experiments*"; four items, $\omega = .72$), and (c) investigative learning (e.g., "*Students are asked to do an investigation to test out their own ideas*"; three items, $\omega = .77$). These aspects play an important role in inquiry-based settings and in fostering scientific literacy, since they relate to different types of opportunities-to-learn in science (Bybee & McCrae, 2011; Bybee, McCrae, & Laurie, 2009). Seidel and Prenzel (2006) provided theoretical underpinnings for the validity of these scales. They argued that being actively involved in classroom discussions about scientific topics (interactivity) positively affects motivational and educational outcomes in general and scientific literacy specifically; and the authors pointed to empirical results supporting this claim (Bransford & Donovan, 2005; Hofstein & Lunetta, 2004). Furthermore, the model assumes that the competence to use scientific evidence – including hands-on activities such as gathering data with experiments – is pivotal for the development of scientific literacy. In addition, they mentioned that "hands-on experiments in the classroom have positive effects on the affective aspects of students' scientific literacy such as interest and attitudes towards science" (Prenzel, Seidel, & Kobarg, 2012, p. 672). Investigative learning goes beyond hands-on activities by including "activities in which the students test their own ideas using scientific methods. Thus, the degrees of freedom are much larger than in hands-on activities" (Prenzel et al., 2012, p. 672).

3.2.4. *Science achievement*

Based on the definition of scientific literacy, which has been used in PISA to develop the achievement tests, the test items measure different scientific competences (Bybee et al., 2009; for details, see OECD, 2007). Three closely related facets of scientific literacy were examined: *Identifying scientific issues* (24 items), *Explaining phenomena scientifically* (53 items), and *Using scientific evidence* (31 items). Students' responses were scored dichotomously or polytomously. By means of multidimensional item response theory, five plausible values were drawn from an ability distribution as measures for each of the three facets of scientific literacy (for details, please refer to OECD, 2009).

In our analysis, we assumed a single latent factor of scientific literacy (global scale), which is consistent with the PISA reports (OECD, 2007; see also Taskinen et al., 2013). Therefore, we specified a CFA model using five sets of three plausible values as indicators to represent the facets of scientific literacy. The reliability of this latent factor was $\omega = .97$.

3.2.5. *Future-oriented motivation*

The student questionnaire contained four items on students' future-oriented motivation. These items reflect the perceived usefulness of science for the students' future vocational life and have previously been used as a proxy for career aspirations and future educational choices (i.e., the motivation to aspire a career in the sciences; Nagengast et al., 2011; Nagengast & Marsh, 2012; Taskinen et al., 2013). Students evaluated their agreement with statements such as "*I will use science in many ways when I am an adult*" on a 4-point scale ranging from 0 (*I strongly disagree*) to 3 (*I strongly agree*). The reliability of the scale was $\omega = .91$.

3.3. *Statistical analyses*

3.3.1. *Structural equation models*

We applied confirmatory factor analysis (CFA) to address the research questions on the relation between students' competence beliefs (Hypotheses 1.1 and 1.2) and to evaluate the reliabilities of the scales. We compared a one-factor model assuming that students' replies to all ASC and ASE items are explained by one common factor with a two-factor model that assumes two factors accounting for interindividual differences in the ASC and ASE items. The robust maximum likelihood estimator (MLR) was used to obtain robust estimations of standard errors and to account for violations of the normality of observations, particularly for categorical items with at least four categories (Flora & Curran, 2004).

Hypotheses 2.1–3.4 were subsequently addressed within a structural equation modeling (SEM) framework. For models involving students' ability, we replicated the analyses for each plausible value (i.e., five times) and combined the resulting test statistics according to Rubin's rules (Enders, 2010), as recommended in the PISA technical report (OECD, 2009, pp. 9–10). In these analyses, we further tested whether ASC and ASE or their antecedents showed differential effects by comparing the model fit of two nested models. In the first model, path coefficients between competence beliefs and the covariates at hand were freely estimated, whereas the second model assumed equal coefficients to indicate no differences in the effects.

In order to investigate the Big-Fish-Little-Pond-Effect (BFLPE; Hypotheses 2.3 and 2.4), we used doubly-latent multilevel structural equation modeling with schools as clusters (Marsh et al., 2009; Parker et al., 2014). We first analyzed contextual effects of aggregated school achievement on students' ASC and ASE in science after controlling for the effect of achievement on the individual level. If the coefficient deviated significantly from zero, the BFLPE was present. Second, we estimated the effect size *ES2* as proposed by Marsh et al. (2009) to evaluate the practical importance of the BFLPE.

3.3.2. *Evaluation of model fit*

Besides the χ^2 value, we evaluated and compared models using the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), and the Standardized Root Mean Square Residual (SRMR). Common cutoff values were applied to evaluate the goodness of model fit (e.g., Hu & Bentler, 1999; Marsh, Hau, & Grayson, 2005): CFI and TLI above .95, RMSEA below .05 or .08, and SRMR values below .06 were regarded as acceptable values. However, little is known about reasonable cut-off values for the SRMR at the cluster level in multilevel modeling. Also note that the χ^2 statistic strongly depends on the sample size and is highly sensitive in large samples, especially in measurement models with many parameters where accumulations of trivial misfit often lead to significant χ^2 values. To compare the fit of concurrent models, the Satorra–Bentler-corrected χ^2 difference test was used (Bryant & Satorra, 2012). However, this test could not be applied to compare models in which plausible values were pooled, since pooled χ^2 values do not necessarily follow a χ^2 distribution in models using the MLR estimator. Instead, we compared differences on the CFI, TLI, RMSEA, and SRMR descriptively. All analyses were conducted in *Mplus 7.11* (Muthén & Muthén, 1998–2013). Due to the large sample size and the high complexity of the models, we chose the 1% level of statistical significance.

3.3.3. *Handling of missing data*

For the scales from the student questionnaire, the proportion of missing values varied between 5.9% and 7.2%. As test performance in science was strongly associated with missingness in students' science self-concept ($d = 1.09$, $SE = .01$) and self-efficacy ($d = 1.13$, $SE = .02$), it appeared reasonable to assume the missing-at-random mechanism. The Full-Information-Maximum-Likelihood method (FIML) and the expectation-maximization algorithm were used to handle missing data because they yield less biased estimates than traditional approaches such as listwise or pairwise deletion while also maintaining statistical power (Enders, 2010).

3.3.4. *Weighting and handling selection bias*

In PISA 2006, students and schools were randomly selected in a two-step procedure. Within this process, the probability of being selected was not uniformly distributed across the federal states of Germany. Consequently, selection biases that could affect the outcomes of maximum likelihood estimations may occur (Asparouhov, 2005). We included sampling weights at the student and school levels and used the TYPE = COMPLEX option in *Mplus* to adjust for the effects of sampling error and the clustering in schools.

Table 1
Descriptive statistics for the scales from the student questionnaire.

Scale	# Items	ω	*M*	*SD*	Min	Max	Correlations between latent variables				
							2.	3.	4.	5.	6.
1. Science self-efficacy	8	.82	14.27	4.58	0	24	.57*	.38*	.10*	.19*	–.01
2. Science self-concept	6	.90	1.35	4.00	0	18		.52*	.26*	.26*	.18*
3. Future-oriented motivation	4	.91	3.84	3.34	0	12			.13*	.18*	.10*
4. Interactive learning	4	.72	6.04	2.77	0	12				.62*	.63*
5. Hands-on activities	4	.72	5.07	2.45	0	12					.72*
6. Investigative learning	3	.77	2.04	1.98	0	9					

Note. # Items: Number of items. ω refers to McDonald's coefficient of scale reliability (McDonald, 1999).
* $p < .01$.

4. Results

4.1. Descriptive statistics

Descriptive statistics for the constructs included in the present investigation are summarized in Table 1. The scales showed acceptable reliabilities and good psychometric properties. No ceiling or floor effects were present. Unidimensional measurement models with sufficiently high goodness-of-fit indexes were established separately for all constructs.

4.2. Measurement separability

To explore the factor structure of students' competence beliefs in science (Hypotheses 1.1 and 1.2), we compared a one-factor model with a two-factor model distinguishing between ASC and ASE. The one-factor model did not fit the data, $SB\text{-}\chi^2$ (77) = 4,253.2, $p < .001$, CFI = .80, TLI = .77, RMSEA = .11, SRMR = .10. By contrast, the two-factor model showed a good fit to the data, $SB\text{-}\chi^2$ (76) = 497.4, $p < .001$, CFI = .98, TLI = .98, RMSEA = .04, SRMR = .02, as expected in Hypothesis 1.1. In this model, the correlation between the two latent factors was moderate ($\rho = .57$, $SE = .02$) and thus supported Hypothesis 1.2. The investigation of the factorial structure provided a clear empirical distinction between the two constructs in the domain of science. However, a distinct measurability is necessary but not sufficient for differential relations to relevant covariates. Such patterns that would underline the practical necessity of a clear-cut dissociation of the two constructs were explored in the following analyses.

4.3. Competence beliefs as outcomes

In order to test our hypotheses on the differential effects of inquiry-related activities in science classrooms on self-concept and self-efficacy (Hypotheses 2.1 and 2.2), we estimated a series of latent regression models (see Table 2). To study the effects of inquiry-related classroom activities, we first estimated the effects on ASC and ASE for each of the three activities – interactive learning, hands-on activities, and investigative learning – separately (see Table 2, Models 1–3). Whereas interactive learning and hands-on activities showed positive effects on both self-beliefs, investigative learning only affected ASC positively. The effects of investigative learning on ASE were not statistically different from zero. In the next step, all three scales representing inquiry-related classroom activities were included as predictors of ASC and ASE in the same regression model (see Table 2, Model 4). Interactive and hands-on science learning showed positive effects on ASC. Also, hands-on activities in science classrooms positively affected students' ASE, whereas interactivity did not contribute. Investigative learning did not longer show substantial relations to ASC when the effects of interactive learning and hands-on activities were estimated in the same model. Moreover, this factor showed a negative effect on self-efficacy. As both direction and magnitude changed for some inquiry-based learning activities dependent on the way of modeling (separately vs. simultaneously including predictor in the regression models), we checked for multicollinearity. The correlation between the three inquiry-related activities ranged from .62 to .72 (see Table 1). Moreover, we calculated the Variance Inflation Factor (VIF; Kline, 2010) for all

Table 2
Results of latent regression models estimating the effects of inquiry-related learning activities on competence beliefs.

	Unconstrained models								Constrained model	
	Model 1		Model 2		Model 3		Model 4		Model 5	
Outcomes Predictors	*ASC* β (*SE*)	*ASE* β (*SE*)	*ASC* β (*SE*)	*ASE* β (*SE*)	*ASC* β (*SE*)	*ASE* β (*SE*)	*ASC* β (*SE*)	*ASE* β (*SE*)	*ASC* β (*SE*)	*ASE* β (*SE*)
Inquiry learning										
Interactive learning	.26 (.02)*	.10 (.02)*					.18 (.03)*	.06 (.04)	.10 (.03)*	.12 (.04)*
Hands-on learning			.26 (.02)*	.19 (.02)*			.19 (.04)*	.37 (.04)*	.26 (.03)*	.31 (.04)*
Investigative learning					.19 (.02)*	.01 (.02)*	–.06 (.04)	–.29 (.05)*	–.17 (.04)*	–.20 (.04)*
Model indices										
R^2 (*SE*)	.07	.01	.07	.04	.04	.00	.09 (.01)	.07 (.01)	.05 (.01)	.07 (.01)
$SB\text{-}\chi^2$ (*df*)	621.6* (131)		755.4* (132)		639.1* (116)		1,586.8 (264)*		1,715.1 (267)*	
CFI	.981		.974		.979		.961		.958	
TLI	.977		.970		.975		.956		.952	
RMSEA	.029		.032		.031		.033		.034	
SRMR	.021		.027		.024		.030		.040	
Model comparison										
$\Delta SB\text{-}\chi^2$ (Δdf)									–	104.4 (3)*

Note. $SB\text{-}\chi^2$: Satorra–Bentler-corrected χ^2 statistic (Bryant & Satorra, 2012).The regression coefficients are standardized. In Model 5, the paths affecting ASE and ASC were constrained to be equal which is reflected in equal unstandardized regression coefficients. The standardized coefficients are unequal, however, due to unequal residual variances of ASE and ASC.
* $p < .01$.

predictors in order to estimate to what extent the variance of a coefficient is "inflated" due to a linear dependence with all other predictors of the model. According to Kline (2010) values above 10 for the VIF indicate problems with multicollinearity. Since the VIFs for all inquiry-related classroom activities were below 2.5 (interactive learning: 2.18; hands-on activities: 2.34, and student investigations: 2.38), we concluded that multicollinearity does not seem problematic in our analyses and that the interpretation of the regression weights is feasible.

In the last step, we estimated a model in which the effects of all three activities on ASC and their effects on ASE were constrained to be equal (see Table 2, Model 5). The model fitted the data worse than the unconstrained model indicating that differential effects of inquiry-based activities on ASC and ASE were present (see Table 2, Model Indices). Hence, we rejected Hypothesis 2.1 on the positive effects of the three science learning scales for investigative learning. As expected, the positive effect of hands-on activities was higher for self-efficacy in support of Hypothesis 2.2.

4.3.1. Effects of average school achievement (BFLPE)

Multilevel models testing this effect for ASC, pooled $SB\text{-}\chi^2$ $(59) = 540.8$, $p < .001$, CFI = .98, TLI = .97, RMSEA = .04, $SRMR_{within} = .03$, $SRMR_{between} = .14$, and ASE in science, pooled $SB\text{-}\chi^2$ $(95) = 600.29$, $p < .001$, CFI = .97, TLI = .97, RMSEA = .03, $SRMR_{within} = .03$, $SRMR_{between} = .07$, fitted the data reasonably well. For self-concept, the BFLPE was strong ($B = -.28$, $SE = .03$, $p < .001$, $ES2 = -.64$). By contrast, no statistical evidence for this effect could be found for students' self-efficacy in science ($B = -.05$, $SE = .03$, $p = .16$, $ES2 = -.12$). Taken together, these findings were consistent with our hypotheses on a negative BFLPE for both constructs (Hypothesis 2.3) and a higher effect on ASC (Hypothesis 2.4).

4.4. Competence beliefs as predictors

Finally, we investigated whether or not students' ASC and ASE would predict their achievement and future-oriented motivation in science and whether these predictive effects would be different (Hypotheses 3.1–3.4). Results are summarized in Table 3.

Using self-concept and self-efficacy as predictors in separate regression models, we obtained significant effects on scientific literacy (Table 3, Models 1 and 2). After including both competence beliefs as related predictors in a subsequent step, only ASE showed a significant effect on achievement, whereas ASC did not explain incremental proportions of variance (Table 3, Model 3). The path coefficients could not be constrained to be equal without a relevant decrease in model fit ($SB\text{-}\chi^2$ $(117) = 1{,}437.7$, $p < .001$, CFI = .963, TLI = .957, RMSEA = .048, SRMR = .080; ΔCFI = −.010, ΔTLI = −.008, ΔRMSEA = +.007, ΔSRMR = +.047; differences in fit indices were pooled across the five plausible values). In conclusion, Hypothesis 3.1 on the positive effects of self-concept and self-efficacy could be retained when estimating separate regression models, although the joint modeling revealed an insignificant path coefficient for self-concept. The stronger effect of ASE on scientific achievement represents conclusive evidence for Hypothesis 3.2.

Next, we studied effects on future-oriented motivation. The results of the separate modeling showed that both competence beliefs positively contributed to future-oriented motivation (Table 3, Models 4 and 5). As expected (Hypothesis 3.3), modeling the effects of ASC and ASE jointly showed that both positive effects remained significant (Table 3, Model 6). In order to test for differences favoring self-concept as a predictor (Hypothesis 3.4), another model assuming equal path coefficients between the two competence beliefs was specified. This model fitted the data worse than the unconstrained model, $SB\text{-}\chi^2$ $(133) = 913.8$, $p < .001$, CFI = .978, TLI = .974, RMSEA = .036, SRMR = .032; $\Delta SB\text{-}\chi^2$ $(1, N = 4{,}891) = 87.7$, $p < .001$; ΔCFI = −.002, ΔTLI = −.003, ΔRMSEA = +.002, ΔSRMR = +.011. Consequently, differential effects were present in support of Hypothesis 3.4.

5. Discussion

ASC and ASE are two related but conceptually different constructs describing students' competence beliefs. In this study, we explored whether their conceptual differences would also translate into an empirical separability by deriving and testing specific theory-based predictions. We looked into the empirical evidence with: (a) a correlational CFA approach focusing on the internal structure of competence beliefs, (b) a (multilevel-) regression approach focusing on the effects of peer ability and inquiry-related classroom activities on competence beliefs, and (c) a regression approach focusing on the predictive validity of competence beliefs on educational outcomes.

5.1. Summary of results

We found that ASC and ASE could be separated in competing measurement models and showed a moderate positive correlation (Hypotheses 1.1 and 1.2 were confirmed). Subsequently, we tested for differential relations with the covariates. As expected, ASC was more strongly affected by the average peer ability than ASE (Hypotheses 2.3 and 2.4 confirmed). Surprisingly, the effect of investigative learning on ASE was negative when the positive effects of interactive learning and hands-on activities were estimated in the same model. However, ASE was more strongly affected by hands-on activities than ASC (Hypotheses 2.1 and 2.2 partially confirmed). With respect to the prediction of desirable educational outcomes, we confirmed that the effect of ASE on students' ability was much stronger

Table 3
Results of latent regression modeling with competence beliefs as predictors.

Predictors	Outcomes					
	Science achievement			Future-oriented motivation (career aspiration)		
	Model 1 β (*SE*)	Model 2 β (*SE*)	Model 3 β (*SE*)	Model 4 β (*SE*)	Model 5 β (*SE*)	Model 6 β (*SE*)
Science self-concept	.28 (.02)*	–	−.04 (.03)	.53 (.01)*	–	.46 (.02)*
Science self-efficacy	–	.53 (.02)*	.55 (.03)*	–	.38 (.02)*	.12 (.02)*
R^2 (*SE*)	.08 (.01)*	.28 (.02)*	.28 (.02)*	.28 (.02)*	.15 (.01)*	.29 (.01)*
Model indices						
$SB\text{-}\chi^2$ (*df*)	457.9*	519.5*	1,055.9*	367.7 (34)*	423.5 (53)*	826.1 (132)*
CFI	.979	.978	.973	.985	.981	.980
TLI	.971	.972	.969	.981	.976	.977
RMSEA	.058	.048	.041	.046	.019	.034
SRMR	.033	.037	.033	.017	.039	.021

Note. $SB\text{-}\chi^2$: Satorra–Bentler corrected χ^2 statistic. The regression coefficients are standardized.
* $p < .01$.

when compared with the effect of ASC, which even disappeared when both competence beliefs were included as predictors in the same model (Hypotheses 3.1 and 3.2 confirmed). However, when exploring the effects on students' future-oriented motivation to pursue a career in the sciences, ASC was the stronger predictor (Hypotheses 3.3 and 3.4 confirmed). As we will show in the following, these results enhance our understanding of competence beliefs and provide implications for the practice of science education.

5.2. Relations to previous evidence and implications for the understanding of competence beliefs

Whereas most studies found moderate positive relations between ASC and ASE, in some studies, correlations were so high that the distinctiveness of the two constructs seemed questionable (Bong et al., 2012; Choi, 2005; Marsh et al., 2004). Using data from a representative German sample and well-established measures, we found a positive correlation of $\rho = .57$, indicating that the two constructs are related but can be clearly separated. The comparatively low correlation was presumably also due to the applied operationalization of ASE with task-specific rather than domain-specific items (Bong & Skaalvik, 2003). The correlation was comparable in size with the correlations between ASC and ASE in math found by Ferla et al. (2009, 2010) who also studied a task-based operationalization of ASE. It was substantially lower than the correlations found by Bong et al. (2012) who used domain-specific operationalizations of ASE. It can be reasoned that the two constructs are distinct when task-based ASE items are used, but that a separation may be more difficult when more general, domain-level ASE items are utilized (Skaalvik & Skaalvik, 2004). Further research using various ASE measures in the same study is needed to study the effects of the different operationalizations.

In subsequent steps of our analyses, we provided evidence that the conceptual definitions of ASC and ASE are sustainable and predict differential relations to antecedents and outcomes: As predicted, the BFLPE was stronger for ASC than ASE. This result strengthens the theoretical argument that academic self-concepts are much more reliant on social comparisons and peer group effects (e.g., effects of ability grouping; Marsh, Köller, & Baumert, 2001) compared with self-efficacy beliefs. As mentioned above, this could be due to negative effects of social comparisons with a high peer standard and positive effects of vicarious experiences with high-ability peers may cancel each other out leading to a practically insignificant effect of peer ability on ASE. Further, we explored differential effects of inquiry-related classroom activities that were used as a proxy for opportunities to obtain mastery experiences. The positive effects of hands-on activities and the marginal negative effect of investigative learning, which were stronger for ASE compared with ASC, are consistent with the conceptualization of ASE as (a) task-specific rather than a general evaluation, and (b) as more strongly affected by mastery experiences. The negative effect of investigative learning on ASE was somewhat surprising and contradicted our expectations. One explanation may be that the items used to measure investigative learning emphasize students' *individual* activities such as designing and developing experiments themselves to test their own hypotheses (e.g., "*Students are given the chance to choose their own investigations*"). Presumably, inquiry-based learning and experimentation in the classroom need to be structured in a way that allows students to have positive rather than negative mastery experiences. If students are faced with a very challenging task such as designing an experiment without having acquired the necessary knowledge and without proper instruction, the chance of failing at that task is high (Areepattamannil, Freeman, & Klinger, 2011), leading to feelings of failure and inadequacy. This in turn could have a negative impact on self-efficacy. Whereas the failure in specific tasks is not problematic in itself and may even facilitate learning (e.g., in the "productive failure" approach of math teaching; Kapur & Bielaczyc, 2012), supporting the students to complete challenging tasks would be important to foster their competence beliefs over a longer period of time. This explanation is consistent with research that has argued that inquiry-based learning can be beneficial but needs to be scaffolded, monitored, and adapted to students' prior knowledge (Hmelo-Silver, Duncan, & Chinn, 2007).

In the third step of our analyses, we found differential effects of the two competence beliefs on educational outcomes. Our result that ASC predicted future-oriented motivation better than ASE is consistent with previous evidence (e.g., Parker et al., 2014). The effects strengthen the theoretical conceptualization of both competence beliefs because frames of reference, which have been shown to affect self-concept more strongly than they affect self-efficacy, play a significant role in career choices. When making the decision to study a certain course at a college or university, students likely compare their achievement and interests across different subjects and choose one of their best subjects (a dimensional comparison affecting ASC but not ASE; Möller & Marsh, 2013). Moreover, ASC is considered more general in nature (Marsh, Roche, Pajares, & Miller, 1997). Because career choices are more likely driven by relative evaluations of general ability in a domain within social and internal frames of reference rather than absolute evaluations of one's ability to master specific tasks, the higher effect of ASC is in line with substantive theories. However, it should be pointed out that prior achievement is also among the most important predictors of career choices. Due to the higher effect of ASE on achievement, there may be additional indirect effects of ASE on career choices.

The distinctly stronger predictive power of ASE on achievement in science, even though it is backed up by previous studies (Ferla et al., 2009; Pajares, 1996; Pajares & Miller, 1994), is somewhat contrary to the idea that ASC and achievement are reciprocally related (e.g., Marsh & Martin, 2011). One explanation might be that ASE items are formulated to match typical tasks that are also part of achievement tests. In the PISA 2006 data, to which our analyses were applied, the ASE items were designed to closely match the topics as well as the task types of the PISA achievement test. Consequently, Marsh et al. (1997) argued that if identical tasks were used for the assessment of self-efficacy and to measure ability, the effect of self-efficacy on achievement would be positively inflated. Similarly, Stankov, Lee, Luo, and Hogan (2012) speculated that the similarity between ASE and test items would moderate the relation between ASE and achievement. If ASE had been measured by tasks that were more different from the PISA achievement test, a lower relation would have been expected. It should further be mentioned that, even though ASE is more highly related to specific test scores, the longitudinal evidence for a reciprocal-effects model between competence beliefs and educational achievement in the long run is more comprehensive for ASC (Marsh, Hau, & Grayson (2005); Marsh & Martin, 2011; (Marsh, Trautwein et al., 2005)). For ASE, the question of causality with regard to achievement has also been discussed but rarely evaluated empirically (Marsh et al., 1997; Pajares, 1996). Such an evaluation would be challenging for ASE, however, because the tasks used to measure self-efficacy are likely not valid indicators across different grade levels because students' abilities, school curricula, and subsequently, typical ASE tasks change over time.

5.3. Limitations

Even though we tested a number of theory-driven hypotheses, we did by no means examine all possible antecedents of self-beliefs or all outcomes that they might affect. Further, some of the relations assumed in the model are likely to be reciprocal (e.g., the relation between competence beliefs and achievement). Thus, we do not claim to test causal effects or to offer a comprehensive model of how motivational factors affect educational outcomes. Rather, we explored differential relations between self-beliefs and selected constructs that represented social comparisons, opportunities-to-learn, and educational outcomes.

In this study, we examined competence beliefs with regard to general science and, hence, measured all available constructs (e.g., achievement, both self-beliefs) for general science. It should be noted, however, that science education in most OECD countries involves several separate subjects. In Germany, these subjects are typically biology, chemistry, and physics. For ASC, it has been shown that German students differentiate between these three subjects in their self-evaluations (Jansen et al., 2014). Such subject-specific measures were not administered in the PISA 2006 study. However, examining science in general is in line with the idea of interdisciplinary scientific literacy as an important educational goal, defined and tested by the PISA studies (Bybee et al., 2009).

5.4. Implications for science education and science teaching in the classroom

In line with previous evidence, our study showed high relations between competence beliefs and desirable educational outcomes. Therefore, the general recommendation to foster students' competence beliefs can be extended and specified, since the two competence beliefs affect two different but desirable outcomes (i.e., science achievement and future-oriented motivation). In addition to other factors (e.g., general cognitive ability), the application of training programs may provide a way to enhance students' educational success in STEM areas. Generally, positive and individual feedback provided by teachers are effective ways to foster both self-beliefs (O'Mara, Marsh, Craven, & Debus, 2006; Schunk, 1991). On the basis of the conceptual differences between ASC and ASE, further specific strategies to promote these competence beliefs may be indicated. For example, ASE can be enhanced by creating opportunities for students to engage in meaningful mastery experiences through the design of classroom activities and tasks (Pajares, 1996). Effective interventions targeted on one of the two constructs do exist for both, ASC (Ginsburg-Block, Rohrbeck, & Fantuzzo, 2006; O'Mara et al., 2006) and ASE (Luzzo, Hasper, Albert, Bibby, & Martinelli, 1999).

Another implication of the differential effects on educational outcomes is that competence beliefs should match the specific outcome to be predicted. When educational aspirations and student motivation are of interest, self-concept serves as a meaningful predictor. When the focus of the study lies on academic achievement, students' self-efficacy is more appropriate. This distinction may be relevant for researchers deciding which constructs to measure, and for practitioners making diagnostic decisions.

Apart from these general implications, the results concerning science teaching may be relevant for classroom practice. It could be shown that students, who reported higher frequencies of hands-on learning activities in their science classes, also showed a higher ASE. This may be an indication that hands-on activities in the sciences provide opportunities for students to gather the mastery experiences that are pivotal for the development of self-efficacy. However, it should be noted that the occurrence of such positive motivational experiences depends on the implementation of these activities and our study only provides a very broad, generalized perspective. Further research is needed to study how the inquiry-based science teaching may be designed to best encourage mastery experiences.

References

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

Areepattamannil, S., Freeman, J. G., & Klinger, D. A. (2011). Influence of motivation, self-beliefs, and instructional practices on science achievement of adolescents in Canada. *Social Psychology in Education*, *14*, 233–259. doi:10.1007/s11218-010-9144-9.

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 411–434. doi:10.1207/s15328007sem1203_4.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Adolescence and education, vol. 5: Self-efficacy and adolescence* (pp. 307–337). Greenwich, CT: Information Age Publishing.

Beghetto, R. A. (2007). Factors associated with middle and secondary students' perceived science competence. *Journal of Research in Science Teaching*, *44*, 800–814. doi:10.1002/tea.20166.

Betz, N. E. (2000). Self-efficacy theory as a basis for career assessment. *Journal of Career Assessment*, *8*, 205–222. doi:10.1177/106907270000800301.

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research*, *105*, 336–352. doi:10.1080/00220671.2011.627401.

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, *34*, 139–153. doi: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. doi:10.1023/A:1021302408382.

Bransford, J. D., & Donovan, M. S. (2005). Scientific inquiry and how people learn. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn – Science in the classroom* (pp. 397–420). Washington, DC: The National Academies Press.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*, 372–398. doi:10.1080/10705511.2012.687671.

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: perspectives from PISA 2006 science. *International Journal of Science Education*, *33*, 7–26. doi:10.1080/09500693.2010.518644.

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: an assessment of scientific literacy. *Journal of Research in Science Teaching*, *46*, 865–883. doi:10.1002/tea.20333.

Chen, J. A., & Pajares, F. (2010). Implicit theories of ability of Grade 6 science students: relation to epistemological beliefs and academic motivation and achievement in science. *Contemporary Educational Psychology*, *35*, 75–87. doi:10.1016/j.cedpsych.2009.10.003.

Choi, N. (2005). Self-efficacy and self-concept as predictors of college students' academic performance. *Psychology in the Schools*, *42*, 197–205. doi:10.1002/pits.20048.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: reconsidering structural relationships. *Learning and Individual Differences*, *19*, 499–505. doi:10.1016/j.lindif.2009.05.004.

Ferla, J., Valcke, M., & Schuyten, G. (2010). Judgments of self-perceived academic competence and their differential impact on students' achievement motivation, learning approach, and academic performance. *European Journal of Psychology of Education*, *25*, 519–536. doi:10.1007/s10212-010-0030-9.

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, *7*, 117–140.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. doi:10.1037/1082-989X.9.4.466.

Freedman, M. P. (1997). Relationship among laboratory instruction, attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching*, *34*, 343–357. doi:10.1002/(SICI)1098-2736(199704)34:4 < 343::AID-TEA5>3.0.CO;2-R.

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, *98*, 732–749. doi:10.1037/0022-0663.98.4.732.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*, 1159–1176. doi:10.1002/tea.20442.

Goetz, T., Haag, L., Lipnevich, A. A., Keller, M., Frenzel, A. C., & Collier, A. P. M. (2014, April). *Thinking impacts feeling. Judgments of school domain similarity and between-domain relations of students' academic emotions.* Paper presented at the Annual Meeting of the American Educational Research Association, Philadelphia, PA, USA.

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

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*, 99–107. doi:10.1080/00461520701263368.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*, 88–140. doi:10.3102/00346543067001088.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: foundations for the twenty-first century. *Science Education*, *88*, 28–54. doi:10.1002/sce.10106.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. doi:10.1080/10705519909540118.
Huang, C. (2011). Self-concept and academic achievement: a meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*, 505–528. doi:10.1016/j.jsp.2011.07.001.
Huguet, P., Dumas, F., Marsh, H., Wheeler, L., Seaton, M., Nezlek, J., et al. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): an integrative study. *Journal of Personality and Social Psychology*, *97*, 156–170. doi:10.1037/a0015558.
Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences*, *30*, 11–21. doi:10.1016/j.lindif.2013.12.003.
Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*, 45–83. doi:10.1080/10508406.2011.591717.
Kelley, T. L. (1927). *Interpretation of educational measurements*. Yonkers-on-Hudson, NY: World Book Company.
Kessels, U., & Hannover, B. (2008). When being a girl matters less: accessibility of gender-related self-knowledge in single-sex and coeducational classes and its impact on students' physics-related self-concept of ability. *British Journal of Educational Psychology*, *78*, 273–289. doi:10.1348/000709907X215938.
Kline, R. B. (2010). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
Luzzo, D. A., Hasper, P., Albert, K. A., Bibby, M. A., & Martinelli, E. A., Jr. (1999). Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students. *Journal of Counseling Psychology*, *46*, 233–243. doi:10.1037/0022-0167.46.2.233.
Marsh, H. W. (1986). Verbal and math self-concepts: an internal/external frame of reference model. *American Educational Research Journal*, *23*, 129–149. doi:10.3102/00028312023001129.
Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*, 280–295. doi:10.1037/0022-0663.79.3.280.
Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: theoretical and empirical justification. *Educational Psychology Review*, *2*, 77–172.
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*(2), 133–163. doi: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. doi:10.1207/S15327906MBR3802_3.
Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: a reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, *96*, 518–522. doi:10.1037/0022-0663.96.3.518.
Marsh, H. W., Hau, K., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275–340). Mahwah, NJ: Erlbaum.
Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, *38*, 321–350. doi:10.3102/00028312038002321.
Marsh, H. W., 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. doi:10.1016/j.learninstruc.2014.04.002.
Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., et al. (2015). Dimensional comparison theory: paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction*, *35*, 16–32. doi:10.1016/j.learninstruc.2014.08.005.
Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., et al. (2009). Doubly-latent models of school contextual effects: integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*, 764–802. doi:10.1080/00273170903333665.
Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: relations and causal ordering. *British Journal of Educational Psychology*, *81*, 59–77. doi:10.1348/000709910X503501.
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. doi:10.1006/ceps.1997.0942.
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. doi:10.1037/0022-0663.100.3.510.
Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416. doi:10.1111/j.1467-8624.2005.00853.x.
Mason, L., Boscolo, P., Tornatora, M. C., & Ronconi, L. (2013). Besides knowledge: a cross-sectional study on the relations between epistemic beliefs, achievement goals, self-beliefs, and achievement in science. *Instructional Science*, *41*, 49–79. doi:10.1007/s11251-012-9210-0.
McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
Möller, J. (2005). Paradoxical effects of praise and criticism: social, dimensional and temporal comparisons. *British Journal of Educational Psychology*, *75*, 275–295. doi:10.1348/000709904X24744.
Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*, 544–560. doi:10.1037/a0032459.
Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus version 7.11* [Computer software]. Los Angeles, CA: Muthén & Muthén.
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. doi:10.1037/a0027697.
Nagengast, B., & Marsh, H. W. (2014). Motivation and engagement in science around the globe: testing measurement invariance with multigroup structural equation models across 57 countries using PISA 2006. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment* (pp. 317–344). Boca Raton, FL: CRC Press.
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. doi:10.1177/0956797611415540.
OECD (2007). *PISA 2006: Science competencies for tomorrow's world: Volume 1: Analysis*. Paris: OECD.
OECD (2009). *PISA 2006 technical report*. Paris: OECD.
Okeke, N. A., Howard, L. C., Kurtz-Costes, B., & Rowley, S. J. (2009). Academic race stereotypes, academic self-concept, and racial centrality in African American youth. *Journal of Black Psychology*, *35*, 366–387. doi:10.1177/0095798409333615.
O'Mara, A. I., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, *41*, 181–206. doi:10.1027/s15326985ep4103_4.
Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education*, *25*, 1049–1079. doi:10.1080/0950069032000032199.
Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*, 543–578. doi:10.3102/00346543066004543.
Pajares, F., Britner, S. L., & Valiante, G. (2000). Relation between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology*, *25*, 406–422. doi:10.1006/ceps.1999.1027.
Pajares, F., & Miller, D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *Journal of Educational Psychology*, *86*, 193–203. doi:10.1037/0022-0663.86.2.193.
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. doi:10.1080/01443410.2013.797339.
Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33–40. doi:10.1037/0022-0663.82.1.33.
Prenzel, M., Seidel, T., & Kobarg, M. (2012). Science teaching and learning: an international comparative perspective. In B. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 667–678). New York, NY: Springer.
Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3–19). New York, NY: Springer.
Scherer, R. (2013). Further evidence on the structural relationship between academic self-concept and self-efficacy: on the effects of domain specificity. *Learning and Individual Differences*, *28*, 9–19. doi:10.1016/j.lindif.2013.09.008.
Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, *26*, 207–231. doi:10.1207/s15326985ep2603&4_2.
Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*, 403–419. doi:10.1037/a0013838.
Seidel, T., & Prenzel, M. (2006). Teaching and learning of science. In ACER (Ed.), *PISA 2006 conceptual framework* (pp. 47–62). Camberwell: ACER. Retrieved from: <http://www.acer.edu.au/files/pisa2006_context_framework.pdf> Accessed 30.06.14.
Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*, 407–441. doi:10.3102/00346543046003407.
Skaalvik, E. M., & Skaalvik, S. (2004). Self-concept and self-efficacy: a test of the internal/external frame of reference model and predictions of subsequent motivation and achievement. *Psychological Reports*, *95*, 1187–1202. doi:10.2466/pr0.95.3f.1187-1202.
Spinath, B., & Spinath, F. M. (2005). Development of self-perceived ability in elementary school: the role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development*, *20*, 190–204.
Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: a better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, *22*, 747–758. doi:10.1016/j.lindif.2012.05.013.
Swann, W. B., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*, 84–94. doi:10.1037/0003-066X.62.2.84.

Taskinen, P. H., Schütte, K., & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: the role of school factors, individual interest, and science self-concept. *Educational Research and Evaluation*, *19*, 717–733. doi:10.1080/13803611.2013.853620.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, *92*, 144–151. doi:10.1037/0022-0663.92.1.144.

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, *98*, 438–456. doi:10.1037/0022-0663.98.2.438.

Tsai, C.-C., Jessie Ho, H. N., Liang, J.-C., & Lin, H.-M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction*, *21*, 757–769. doi:10.1016/j.learninstruc.2011.05.002.

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: a validation study. *Contemporary Educational Psychology*, *34*, 89–101. doi:10.1016/j.cedpsych.2008.09.002.

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

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

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*, 845–862. doi:10.3102/00028312031004845.