Deckblatt für die Dissertation

The Developmental Interplay of Academic Self-Concept and Achievement and the Role of Potential Moderators

---

Inaugural-Dissertation

in der Fakultät Humanwissenschaften

der Otto-Friedrich-Universität Bamberg

vorgelget von

Daniel Sewasew

Bamberg, den 17.04.2019

Tag der mündlichen Prüfung: 10.07.2019

Dekan/Dekanin: Universitätsprofessor/- in Prof. Dr. Jörg Wolstein

Erstgutachter/-in: Universitätsprofessor/- in in Prof. Dr. Ulrich Schroeders

Zweitgutachter/-in: Universitätsprofessor/- in Prof. Dr. Maximillian Pfost

Funding Statement

This work was supported by the Bamberg Graduate School of Social Sciences which is funded by the German Research Foundation (DFG) under the German Excellence Imitative (GSC1024).

URN: urn:nbn:de:bvb:473-opus4-552116
DOI: https://doi.org/10.20378/irbo-55211

Acknowledgments

After an intensive period of four years, today is the day writing this note of thanks is the finishing touch on my dissertation. It has been a period of intense learning for me, not only in the scientific arena but also on a personal level. Writing this dissertation has had a big impact on me. I would like to reflect on the people who have supported and helped me so much throughout this period. First I would like to begin with thanking the members of the examining committee (Prof. Dr. Jörg Wolstein, Prof. Dr. Ulrich Schroeders, Prof. Dr. Maximillian Pfost) for agreeing to serve on my committee and lending your expertise.

I would like to express my earnest gratitude to my advisor Prof. Dr. Ulrich Schroeders for the unceasing support to my Ph.D. study and research: for his tolerance, inspiration, keenness, and immense knowledge. During our course of interaction in the last years, I have learned extensively from him, including how to raise new possibilities, how to regard an old question from a new perspective, how to approach a problem by systematic thinking, data-driven decision making and exploiting serendipity. Thank you so much for making this experience meaningful and for challenging me to do my best. I could not have imagined having a better supervisor and mentor for my Ph.D. study.

My sincere thanks also goes out to Prof. Dr. Cordula Artelt for her insightful comments and encouragements, but also for the valuable questions which incented me to widen my research from various perspectives during the initial stage of the first paper.

Also, I am grateful to Dr. Irene M. Schiefer, in particular for enlightening me the first glance about BiKS dataset and her great contribution throughout the first paper.

A special thanks to Dr. Sebastian Weirich, for his tireless efforts in analyzing the psychometric properties of the BiKS data to a get better version out of it. Without his support and feedback, it would not have been possible to complete the first paper.

My special gratitude also goes out to Dr. Lynne Koester for her amazing support and collaboration since the application to pursue my Ph.D. study, and her contribution as the co-author for the second paper was extremely valuable.

In my daily work, I have been blessed with a friendly and cheerful group of fellow Ph.D. candidates in the graduate school. Ashenafi, Dorothea and her husband Oliver, Marc, Simon, Ansgar, and Elena—were not only able to support each other by deliberating over our problems and findings but also happily by talking things other than just our papers.

Very special gratitude goes out to all down at the BAGSS administrative staff— Prof. Dr. Thomas Saalfeld, Ms. Miriam Schneider (former staff), Katrin Bernsdorff, Olga Rayevska, and the whole circle who was always ready to give their timely help whenever required. And, I must thank the research funder (Deutschland Research Foundation, DFG) for helping and providing the funding for the work.

Last but not least, I would like to thank my family: my parents and to my sisters and brothers for supporting me spiritually throughout writing this dissertation and my life in general. I would like to extend my appreciation to Mimi and Elias for making my stay in Bamberg easy and enjoyable; I never forget the time spent with them usually on the weekends. Above all, I owe it all to Almighty God for granting me the wisdom, health, and strength to undertake this research task and enabling me to its completion. Equally, I want to thank me for doing all this hard work.

Table of Content

Abstract

The development of positive academic motivation and high achievement has a sweeping impact on students' desirable outcomes (e.g., academic successes, occupational aspirations, and lifelong achievements). This dissertation sought to investigate the complex developmental relationships of academic motivations (i.e., verbal and math self-concepts) and academic achievements (i.e., verbal and math competencies), and the role of potential moderators (i.e., ethnic-background and sex) in primary and secondary school-aged children. The dissertation utilized the U.S. and Germany sample of students who participated in the large-scale longitudinal data collection and applied cross-lagged panel analyses. In line with the prediction of the reciprocal effects model (REM), the results highlight positive developmental relations between academic self-concept and achievement within a domain in primary and secondary school-aged students. Whereas, consistent with the presumption of Internal/External frame of reference model (I/EM) negative relations across two non-matching domains (i.e., verbal and math) were found among primary-aged students. Besides, the effect of prior achievement on self-concept was much stronger than the effect of self-concept on achievement in primary school-aged students, whilst the reverse was stronger in secondary school-aged students in both high and low-stake measures (i.e., standardized test-scores and grades). Therefore, there is a pattern of mutual effects that hinting at the domain specificity of these constructs and the multidimensionality of academic self-concepts, for which the factor analyses provided strong support as well. Moreover, the dissertation also highlights the non-generalizability of REM across ethnic-groups, i.e., the ethnic-background is a potential moderator of the reciprocal effects between academic self-concept and achievement (i.e., verbal) in primary school-aged children. However, this dissertation also provided evidence that sex is not a potential moderator of the reciprocal effects between academic self-concept and

achievement (i.e., math) in secondary school-aged students. In conclusion, the dissertation accentuated the importance of improving academic motivation and achievement jointly. Results stress the need for further investigation of the underlying causal mechanisms between academic self-concept and achievement development, and for an integration of the major models in academic self-concept formation into a unifying theoretical framework (i.e., across domains, time and school/class level in primary and secondary school-aged children).

Introduction

The positive self-beliefs centering around how normal, healthy, and exceptional people can acquire the maximum from a lifetime are at the core of the positive psychology revolution (Marsh & Craven, 2006; Seligman & Csikszentmihalyi, 2000). This positive psychology initiative, with its focus on academic motivation [1] (such as academic self-concept), draws enormous attention in the fields of education and psychology (Marsh et al., 2017; Weinert, 1999). Self-concept is “a construct of pervasive significance” (Craven & Marsh, 2008, p.1) that helps the realization of necessary psychological, behavioral, and educational outcomes that underpin human potential. Per se actions (e.g., interventions) designed to improve self-concept will also enhance psychological wellbeing and impact desirable psychological and educational outcomes, whilst those that neglect self-concept may result in negative impacts on these variables (Craven & Marsh, 2008).

Looking at the scientific discourse of self-concept studies, two perspectives dominate the debate. In initial accounts, self-concept did not refer to any specific domain (i.e., unidimensional construct, Rosenberg, 1965). In later records, multidimensional natures of self-concept were more vividly evidenced in many investigations. That is, self-concept is theorized as a multidimensional and hierarchically ordered structure (Shavelson, Hubner, & Stanton, 1976). At the apex of the structure, there is a global self-concept—an individual global self-perception. Further, it has two facets: academic and nonacademic. The academic self-concept further branches into domain-specific segments: Math, English, History, and Science self-concept. The nonacademic self-concept part splits into social, emotional, and physical self-concepts (Shavelson et al., 1976).

[1] In this dissertation, *academic motivation* refers to verbal and math self-concept whereas *academic achievement* verbal and math competence.

There have been laudable attempts to investigate the relations of academic self-concept and academic achievement, and their relations validated across academic domains, age, educational system and culture (Chen, Yeh, Hwang, & Lin, 2013; Guo, Marsh, Parker, & Dicke, 2018; Marsh et al., 2015; Valentine & DuBois, 2005). Also, the developmental aspect of academic self-concept and achievement relations has been studied intensively, such as in a) the reciprocal effect model (REM) which shows that prior academic self-concept is reciprocally related to subsequent achievement and vice versa, and that students make *temporal comparisons*—comparing their current performance with previous competence in the same domain (*external comparisons,* Marsh et al., 2005; Möller, Retelsdorf, Köller, & Marsh, 2011); b) the internal/external frame of reference model (I/EM), in which academic achievement is negatively associated with the non-corresponding self-concept, students make *dimensional comparisons*—comparing their own achievements in different academic domains (*internal comparisons*), and academic achievement is positively associated with the corresponding self-concept, and *social comparisons*—students compare themselves with others (*external comparisons*, Marsh, 1986; Möller, Pohlmann, Köller, & Marsh, 2009); and c) the reciprocal internal/external frame of reference model (RI/EM) indicating positive and negative developmental relations between academic self-concepts and achievements within and across two academic domains (Möller & Marsh, 2013; Möller et al., 2011). The models have been tested based on the cross-sectional and longitudinal study designs to investigate the complex developmental relationships of academic self-concept and achievement. Overall, studies revealed support for the REM (e.g., Marsh et al., 2005), the I/EM (e.g., Möller et al., 2009), and the RI/EM models (e.g., Möller, Zimmermann, & Köller, 2014; Niepel, Brunner, & Preckel, 2014).

However, despite the strong practical support for the relations between academic self-concept and achievement, rare studies have investigated the development of achievement-motivation in multiple content domains (usually verbal and math, but possibly more than two) from the beginning of primary school until grade 5. Instead, most of the existing achievement-motivation studies have relied largely on secondary school-aged children, even though the primary school might be the most important period in the establishment of children's strong achievement-motivation (Guo et al., 2018) and fast progress in cognitive and academic skills (Harter, 2012).

In addition, while empirical studies have obviously established the importance of temporal, social and dimensional comparisons for academic self-concept formation, very few studies have empirically investigated their joint influence in primary school-aged children (see Wolff et al., 2018, for a recent exception). As such, in spite of a plethora of secondary school investigations, most of the individual studies used only one of the achievement indicators (either test scores or school grades), essentially none have exclusively juxtaposed the two over a period of time to give a full picture in relation to the developmental point of view. Moreover, there is sparse research on the moderating role of ethnic-background, sex and socioeconomic status (SES) of the student's family in relation to the developmental perspective of achievement-motivation. Crucially, the existing studies have not taken into account the methodological standards (e.g., ensuring longitudinal and multi-group measurement invariance testing, specifying achievement measure at the latent level using item response theory), that are the current state-of-the-art and preconditions in longitudinal studies.

In this context, the present dissertation aimed to rigorously test: a) the developmental interplay between academic self-concept and a standardized achievement measure within and across two domains (i.e., verbal and math) in primary school-aged children; b) the reciprocal relations between academic self-concept and achievement from grade 1 to 5; c) the developmental relationships of academic self-concept and achievement in secondary school sample students using high and low-stake achievement measures; d) the role of potential moderators (i.e., ethnic-background and sex) in affecting the developmental relations of achievement-motivation in primary and secondary school, respectively. To undertake the aforementioned goals, this dissertation is organized in the following sections. The first section focused on the theoretical and empirical issues, the second section deals with the research questions and summarizes the results. The discussion is in the third and last section presented the theoretical and practical implications of the main findings of the studies in addition to their major limitations and potential future research directions.

Theoretical Concepts and Empirical Examinations

Background

The concept of self-concept was first introduced by Carl Rogers and Abraham Maslow, addressing how people perceive themselves as pertaining to their capabilities, attitude, values, and uniqueness (Pastorin & Doyle-Portillo, 2013). Subsequently, the establishment of *Self-Perception Theory* (i.e., individuals use their overt behavior to make inferences about or justify their inner feelings, Bem,1972), paved a way for the emergence of several competing theories of self-perception in the area. Researchers have offered many models and ideas in relation to students' perceptions of their age-related academic abilities and experiences. With regards to students' academic learning, a wide range of concepts (e.g., self-efficacy, self-concept) are used to elucidate students' self-perceptions of their performances associated with their academic achievement. Self-efficacy refers to domain-general self-concept which reveals a student's evaluation of his or her academic competence in different subjects (e.g., "I am good at most school subjects"); whereas self-concept is more domain-specific, reflecting student's judgement of his her ability in a specific academic subject area, such as mathematics/ English (e.g., "I am good at mathematics/English", Craven & Marsh, 2008). However, in the present discussion, domain-specific academic self-concept is the primary focus. In educational psychology, self-concept is conceptualized as an individuals' general perceived ability and motivation to learn and stand out academically (Guay, Marsh, & Boivin, 2003; Marsh et al., 2005). It also describes the students' subjective beliefs concerning their strengths and weaknesses in different school subjects (Wolff, Helm, et al., 2018).

Shavelson et al. (1976) theorized that academic self-concept is multilayered, suggesting students may perhaps recognize competence for particular domains. That is, students' self-concept could differ relative to specific subject domains, for instance, math self-concept could be defined as the students' confidence that they can succeed in math, and verbal self-concept referred to learners beliefs that they can excel in reading performance (Marsh & Martin, 2011; Marsh & Redmayne, 1994). In the present dissertation, borrowed from Marsh et al. (2005) academic self-concept refers to the student's perceived academic competence in specific subjects (e.g., verbal or math). On the other hand, academic achievement literally refers to the student's level of competence, control, or self-confidence in a particular domain (Schiefele, Schaffner, Möller, & Wigfield, 2012).

Success in educational attainment is closely tied to academic self-concept, and its formation is basically rooted in social comparisons, i.e., students use the achievement of relevant others (in school, usually their classmates) as a frame of reference to evaluate or judge their own achievement level (Seaton, Marsh, & Craven, 2010). A plethora of literature has addressed the relationship between academic self-concept and achievement. Studies have elucidated the substantial influence of academic self-concepts on achievement (Huang, 2011; Marsh et al., 2005; Valentine, DuBois, & Cooper, 2004). For example, when students are self-confident in their ability to be academically effective, their self-concept will impact their interest, making them happier about learning. At the point when children demonstrate premium or fervor for learning they will probably turn out to be internally motivated to learn, which will prompt them to push for objectives of scholastic excellence (Frenzel, Pekrun, & Goetz, 2007; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014; Rittmayer & Beier, 2009). For teachers and parents, the establishment of a strong academic self-concept should be the first priority over the promotion of long-term academic achievement (Marsh et al., 2005). Knowledge of students' perceptions of their own academic competencies, feelings, and experiences is significant, for these perceptions will impact how they in turn value academics. In addition, students' perceptions of the significance of academics will eventually influence how well or ineffectively they perform academically (Peterson & Miller, 2004).

To this end, the overall achievement-motivation relations are highlighted; however, it is imperative to understand the complex developmental relationships between academic self-concept and achievement. Hence, in the upcoming sections of this dissertation, first, three major theoretical explanations (i.e., REM, I/EM, and RI/EM) between academic self-concept and achievement will be discussed with their empirical evidence. Second, relevant domain-specific moderators (ethnic-background and sex of students) that affect the relationships of academic self-concept and achievement will be addressed in different school subjects (verbal and math). Finally, the research gaps in the present thesis will be identified.

Within-Domain Developmental Relations Between Academic Self-Concept and Academic Achievement

Reciprocal effect model: The REM is a reconciliation of the two classical opposing "either or" views. That is, either earlier achievement determines later academic self-concept via social comparison process (*skill development model*) or prior academic self-concept determines subsequent achievement directly or via academic choice behavior, higher aspirations, effort, and investment (*self-enhancement model*). However, the REM merged the two models and claimed that earlier academic achievement affects later academic self-concept, and in the same vein, prior academic self-concept affects subsequent academic achievement (Marsh & Craven, 2006). An abundance of studies have reported the close relationship between academic self-concept and achievement for primary education (e.g., Guay et al., 2003; Helmke & Van Aken, G, 1995), and for secondary school education (e.g., Marsh et al., 2005; Retelsdorf, Köller, & Möller, 2014); for a complete overview see (Huang, 2011; Marsh & Martin, 2011; Valentine, DuBois, & Cooper, 2004). Overall, studies reveal that the REMs find strong support when the corresponding academic self-concept and achievement is domain specific (e.g., math or verbal), and when educational progress is determined by grades and teachers' feedback (Huang, 2011; Valentine et al., 2004).

However, despite studies increasingly endorsing REM, the results concerning the direction, strength, and significance have been mixed. Apparently, although totally balanced academic self-concept and competence relations are desired, these are seldom found in the literature. In primary education, reciprocal relations have been obtained inconsistently, especially for very young children (see, Chapman & Tunmer, 1997; Skaalvik & Valås, 1999), which could be partly attributed to the prematurity of children's self-concept, as academic self-concept become more firmly established and stable with age (Chen et al., 2013). In secondary education, most individual studies have included only one of the indicators of achievement (i.e., mostly grade, and seldom test scores), and actually none have juxtaposed the two in relation to developmental perspective at a latent level using longitudinal data in a heterogeneous sample of secondary school students (Marsh et al., 2017; Sewasew, Schroeders, Schiefer, Weirich, & Artelt, 2018). Overall, investigation of the REM with rigorous statistical analysis—particularly applying the current state of the art i.e., using item response theory for achievement measure and maintaining longitudinal and multi-group measurement invariance testing are barely found in the literature.

Across-Domain Developmental Relations Between Academic Self-Concept and Academic Achievement

Internal/External frame of reference model: The development of academic self-concept is grounded on a number of judgment procedures comparing a certain target with a certain standard (Wolff, Nagy, Helm, & Möller, 2018). For instance, the I/EM hypothesizes that students form their self-concept in an academic domain (e.g., math or verbal) by comparing their own achievement (target) concurrently to an external standard (e.g., the achievement of their classroom friends, *social comparison*, Festinger, 1954) along with to an internal standard (e.g., their own achievement in other domains, *dimensional comparison*, Möller & Marsh, 2013). Using a developmental perspective of I/EM assumptions, negative cross-domain effects were found between verbal/math achievement on contrasting subsequent academic self-concepts (*internal frame of reference effects*), and positive within-domain effects were found between verbal/math achievement and the corresponding subsequent self-concept (*external frame of reference effects*).

In one of the most comprehensive meta-analyses (Möller et al., 2009), integrated the results of 68 data sets with more than 125,000 participants providing strong evidence for the I/EM. Concerning the achievement measures, I/EM relations were found both when achievements (i.e., math and verbal) were measured with grades, and when this was done with standardized test results (Möller et al., 2009; Wolff, Nagy, et al., 2018). In addition there has been growing support for predictions based on the I/EM by many different methodological approaches: experimental studies (Möller & Köller, 2001; Müller-Kalthoff et al., 2017; Pohlmann & Möller, 2009; Wolff, Helm, et al., 2018); cross-sectional (Lohbeck & Möller, 2017; Marsh & Hau, 2004; Pinxten et al., 2015); and longitudinal field studies (Chen et al., 2013;

Möller et al., 2011; Möller, Zimmermann, & Köller, 2014; Niepel et al., 2014; Wolff et al., 2018). Moreover, the I/EM has been extended to different domains than math and verbal (Jansen, Schroeders, Lüdtke, & Marsh, 2015; Möller, Streblow, Pohlmann, & Köller, 2006). Notwithstanding this large body of research, with few exceptions (e.g., Wolff, Helm, et al., 2018) the I/E has rarely been examined in primary school longitudinally; these exceptions have predominantly been ones that have also estimated the reciprocal relations of academic self-concept and achievement.

Within-and-Across-Domain Developmental Relations Between Academic Self-Concept and Achievement

Reciprocal internal/external model: This model is the latest one, which unifies dimensional comparison (i.e., students compare their performance across two different domains, usually math and verbal), temporal comparison (i.e., students relate their current performance to prior performance in the same domain), and social comparison (i.e., students evaluate their performance in comparison to others). For example, students develop a higher academic self-concept if they judge their achievement superior in comparison (i.e., down comparison from a better-off target to a worse-off standard) with their classmates (social comparison, Festinger, 1954), with their prior achievement (temporal comparison, Albert, 1977), and with their achievement in other subjects (dimensional comparison, Möller & Marsh, 2013), and vice versa (Wolff et al., 2018). Integrating the I/EM and the REM complements the comparisons of each individual model and gives a full picture of the underlying processes (Marsh & Köller, 2004): the REM lacks the cross-domain perspective, while the I/EM disregards the developmental aspect and the self-enhancement effects. Remarkably, the RI/EM has only been tested with secondary school students in European countries. For German students, studies reported among others positive reciprocal effects of academic self-concept and achievement (grades) within a domain and negative effects of achievement on subsequent self-concepts across domains (see Möller et al., 2011; Niepel et al., 2014). Similarly, Möller et al. (2014) found positive longitudinal effects of achievement and self-concept within domains using grades and test-scores and negative effects of achievement on subsequent academic self-concept across domains. However, taking into account prior achievement the effects of academic self-concept on subsequent achievement across domains were near zero. In a sample of Taiwan students, Chen et al. (2013) longitudinally studied two cohorts of secondary school students and found reciprocal relations between math and Chinese. However, no negative cross-domain effects from prior achievement to subsequent academic self-concept were present.

A recent longitudinal study with primary school students (grade 4 and 5) in Germany (Wolff et al., 2018), replicated the typical pattern of I/EM results: strong positive paths from achievement (grades) to matching self-concepts (social comparison process) and moderate negative paths from achievement to non-matching self-concepts (dimensional comparison process). Moreover, in the long run, the authors found small positive effects from achievement to matching self-concepts (showing temporal evaluation processes within the subjects), and non-significant effects to non-matching self-concepts (signifying that temporal comparison processes within the one domain do not affect self-concept formation in the other domain). However, their study was confined to the use of grades as the achievement measure, and the generalization of the RI/EM to test scores as achievement indicator is still pending in a primary school setting.

Overall, a differing relation between academic self-concept and competence for primary and secondary school students is confirmed by a wide review of the literature. A number of reasons could be forwarded for these diverging results: a) different operationalization of academic achievement (grades vs. test-scores); b) the breadth of the definition of academic self-concept, for example, academic enjoyment and competence beliefs (Else-Quest, Hyde, & Linn, 2010; Pinxten et al., 2014), as well as academic confidence (Else-Quest et al., 2010; Ganley & Lubienski, 2016) as part of self-concept; c) different methodological approaches, that is linear regression with manifest indicators vs. latent variable modeling (Marsh et al., 2005; von Maurice, Dörfler, & Artelt, 2014); d) the average ability level of the sample (e.g., academic track only, see also (Marsh et al., 2017); and e) design of the study, that is, cross-sectional vs. longitudinal (Else-Quest et al., 2010; Marsh et al., 2005).

The Role of Moderators in the Relationships of Academic Self-Concept and Academic Achievement

While an increasing body of investigation has tested the relations between academic motivation and achievement, there is sparse research on the moderating role of ethnic-background, sex and SES in the relations of domain-specific academic self-concept and achievement (e.g., verbal and math). Theoretically, a moderator (e.g., sex or ethnic-background) could affect the direction and strength of the relationship between academic self-concept and achievement. Studies also evidenced that the socioeconomic, cultural, and psychological situations of students affected the formation of achievement-motivation during their primary and secondary school years. Thereby, beyond the longitudinal interplay of academic self-concept and achievement, it is vital to examine the effects of ethnic-grouping and sex on students' academic self-concept and achievement. For example, motivational outcomes (i.e., academic self-concept) have been suggested by some as one explanatory reason for ethnic disparities in academic achievement (Ogbu, 1993); however, differences in academic motivation across groups do not completely reveal inequalities in academic achievement. Despite the fact that academic self-concept of immigrant students tends to be similar or even higher than that of native students, immigrant students often show lower academic competence in comparison to native students (see Kigel, McElvany, & Becker, 2015; Miyamoto et al., 2017, for an overview).

Such conflicting circumstances of academic self-concept and performance (e.g., verbal) could be explained by the aspiration-achievement paradox approach. On the one hand, the *immigrant optimism* approach hypothesizes that immigrant parents leave their home countries due to economic reasons, but then find themselves at society's lowest rank in the host country (Kao & Tienda, 1995; Kigel et al., 2015). Nonetheless, they expect their children to be successful, as educational achievements are perceived to be the key to upward mobility (Mickelson, 1990). Consequently, immigrant parents are disposed to have considerably greater academic hope and expectations for their children; this is often accompanied by children's positive learning attitudes such as high levels of reading motivation (Villiger, Wandeler, & Niggli, 2014). On the other hand, the *stereotype threat* theory suggests that students may experience negative achievement outcomes (e.g., lower standardized test scores and less engagement with academics) because they are burdened by the view of confirming stereotypes impugning their intellectual and academic abilities (Steele, Spencer, & Aronson, 2002).

A large volume of research has examined the gap in performance between ethnic groups; for instance, in the U.S. and Germany White/native students score higher on average on verbal and math tests than all other ethnic groups, particularly when compared to Black and Hispanic/nonimmigrant students (e.g., Bécares & Priest, 2015; Miyamoto et al., 2017). Although few studies have investigated the moderating role of migration background on the developmental relation of verbal achievement-motivations, their results are rather contradictory. For example, the reciprocal relations between verbal achievement and self-concept are not restricted to whether students have an immigration background or not (see Möller et al., 2014); conversely, reciprocal relations are only evidenced in native secondary school students (see Miyamoto et al., 2017; Schaffner, Philipp, & Schiefele, 2016).

Another important moderator that affects the relationship between academic self-concept and achievement is the sex of students. Investigating the generalizability of achievement over sex, especially in sex-stereotypic subjects such as math and verbal, can help to understand the reasons for a gender gap in STEM subjects (Halpern et al., 2007). Marsh theorized and tested a *differential socialization hypothesis* in which “sex-linked differences in socialization patterns may fail to reinforce adequately boys’ positive attitudes, expectations, and performance in verbal areas as well as fail to reinforce adequately girls’ positive attitude, expectations, self-concepts, and performance in mathematics” (Marsh, 1989, p. 195). Differential reinforcement directly refers to differential relations in the REM. Based on Marsh’s idea, it is expected that the link between math self-concept and achievement could be stronger for boys than for girls. However, only a few studies have examined this sex-specific assumption. In a longitudinal study among students attending grade 6 and 7, reciprocal causal relations among math self-concept, interest, and academic achievement were somewhat moderated by the sex and the schooling context (Yoon et al., 1996); that is, math self-concept positively affected achievement for males, whereas the relation was negative for females. More recent studies also reported small sex differences (Pinxten et al., 2014), and the strength of effects from math performance to interest may be slightly stronger for girls than for boys (Ganley & Lubienski, 2016). However, on a meta-analytical level, no sex differences were found in the reciprocal relations (Huang, 2011; Valentine, DuBois, & Cooper, 2004). A similar sex-invariant outcome was reported when academic achievement, self-concept, and interest were juxtaposed (see Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014). In the next subsection, the central research gaps that provided inspiration for the investigations in this dissertation are discussed.

Study Gaps

Empirical support was evidenced by the strong relationship between academic self-concept and achievement, which is captured in a wide range of theories. For example, the RI/EM model postulates positive and negative developmental relations between academic self-concepts and achievements within and across two academic domains. However, the empirical investigations of the RI/EM have been focused on secondary school students and specific countries (e.g., Belgium, Germany, and Taiwan). And what remains an open question and still unclear is, does the RI/EM also apply to other educational settings such as primary schools and other countries?

At first glance, it seems that studies are increasingly endorsing the developmental relation of verbal competence and self-concept (for an overview, see Möller, Retelsdorf, Köller, & Marsh, 2011; Möller, Zimmermann, & Köller, 2014; Niepel, Brunner, & Preckel, 2014; Retelsdorf et al., 2014). Nevertheless, a closer overview of studies reveals an imbalance in the strength and direction of path coefficients, thereby the relationships are inconclusive. In addition, given that the achievement-motivation relations are domain specific (Marsh & Craven, 2006), and the existing studies primarily rely on a certain domain (i.e., Math subject) and ability samples (i.e., secondary school-aged students), the generalizability across subdomains (e.g., reading competence and self-concept, Retelsdorf et al., 2014) and primary school samples, is still unsettled. Also, and particularly in primary school settings, studies have often neglected the potential long-term effects of cultural differences in the early years of schooling. Moreover, the role of relevant moderators (such as ethnic-background) has been overlooked in the developmental relations between verbal self-concept and achievement.

Fortunately, small sex differences in math achievement in favor of boys have frequently been reported in secondary school students (Brunner et al., 2013), but they are not the major factor explaining differences in the enrollment for STEM subjects. In this context, researchers have emphasized the importance of sex-related differences in motivational constructs such as domain-specific self-concept and interest in course selection and educational choices (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). Based on a wide literature review, very little is known regarding whether sex differences in math achievement result from the disparity in the dynamics of self-concept or *vice versa*. This is especially unfortunate since such knowledge would be valuable for the theoretical underpinnings of educational interventions that aim to close the gender gap. In addition, the available REM studies neglect the mean level changes between succeeding measurement time points, that is, they are restricted to the analysis of cross-lagged (i.e., the relation between constructs overtime) and stability (i.e., the relationships within the constructs over time) relations (e.g., see Ganley & Lubienski, 2016; Pinxten et al., 2014). Consequently, reconsidering the prospect of causal order between math achievement and self-concept has intriguing theoretical and practical implications. That offers different tangible answers in developing self-concept interventions to resolve learning concerns (e.g., underachieving students) via improving motivation abilities (e.g., maximizing students' fascination for STEM subjects). An intervention to boost academic achievement over improving self-concept would be redundant if academic self-concept could not be evidenced to benefit academic achievement (O'Mara, Marsh, Craven, & Debus, 2006). Given the above study gaps, in the following section, the core specific research questions and associated results of the three studies of this dissertation will be presented.

Research Questions and Summary of Results

The central aim of this dissertation is to investigate the complex developmental relations between academic motivations (i.e., verbal and math self-concepts) and academic achievements (i.e., verbal and math competencies) in primary and secondary school-aged children. More specifically, the first study (Sewasew & Schroeders, 2019) examined the developmental interplay between self-concept and a standardized achievement measure within and across two domains (i.e., verbal and math) in primary school students. The second study (Sewasew & Koester, 2019) investigated the reciprocal relations between verbal self-concept and competence and the moderation role of students' ethnic-background from grade 1 to 5. The last study (Sewasew et al., 2018) focused on the development of sex differences in the reciprocal relations of academic self-concept and achievement in secondary school sample students using high-and low-stake achievement measures. In the present section of the dissertation, the summary of research questions from each paper and subsequent results will be presented.

Study I. The Developmental Interplay of Academic Self-Concept and Achievement Within and Across Domains Among Primary School Students

Study one has two central objectives. The first one examined the generalizability of REM within the RI/EM over within-domain between academic achievements and motivations (i.e., verbal and math); that is, whether there were positive effects from earlier achievement (verbal and math) on later self-concept within a given domain (*skill development effects*) and positive effects from earlier academic self-concept to later achievement in matching domains (*self-enhancement effects*). The second study investigated the generalizability of I/EM in the RI/EM within and across academic domains between achievements and motivations (i.e., verbal and math) over time; that is, whether there were positive within-domain effects between verbal/math achievement and the corresponding self-concept (*external frame of reference effects*), and negative cross-domain effects between verbal/math achievement on contrasting subsequent academic self-concepts (*internal frame of reference effects*). In this dissertation, a representative longitudinal data set from the U.S. (Early Childhood Longitudinal Study-Kindergarten, ECLS-K) from grade 1 to 5 was used to test the aforementioned research questions. To make the present analyses transparent and reproducible (Nosek et al., 2015), I have provided all syntax online at the OSF (https://osf.io/tdnfa/, Center for Open Science, 2018).

Before the main analyses, the longitudinal measurement invariance (MI) across time points (grades 1, 3, and 5) was tested, which is a necessary prerequisite for making valid statements based on longitudinal data. Subsequently, the model with the most restrictive constraints (strict MI) that was supported by the data was taken as the baseline model to test the RI/EM. *The REM part of the RI/EM*: For the verbal and the math domain, reciprocal relations were evidenced between third and fifth grade (for detail account please see appendix 1, result section). That is, prior academic achievement positively and significantly predicted the individual differences in subsequent academic self-concept beyond the effect of prior self-concept (skill-development effects) for both time periods, and for both language and math. Also, early academic self-concept positively and significantly predicted the individual differences in later academic achievement over and above the effect of prior academic achievement (self-enhancement effects) for language and math. In both domains, the effect of prior academic achievement on subsequent academic self-concept was stronger than in the opposite direction. In terms of underlying concepts (Calsyn & Kenny, 1977), the skill-development paths were more pronounced than the self-enhancement paths.

The I/E part of the RI/EM: The results provided strong support for the internal/external frame of reference effects at both time periods (please see appendix 1, result section). That is, the internal frame of reference effect was observed cross-sectionally—for language on math, and for math on language. Similarly, a typical external frame of reference effect was observed at a given time point, for language and math. In line with the cross-sectional results, we also found longitudinal RI/E effects, that is, negative cross-domain effects of prior achievement on subsequent self-concept—for language on math, and for math on language. Largely, in terms of the underlying concepts (Möller et al., 2009, 2011), the assumptions of the internal/external frame of reference—that is, dimensional and temporal comparisons—were replicated longitudinally among primary school students. Overall, the findings of the study strongly supported the RI/EM for primary school students. The results are compared to previous findings in the literature for secondary school students and discussed with regard to self-concept formation in primary school.

It is important to note, the analyses further revealed close-to-zero correlations between prior verbal self-concept and subsequent math competence between Grade 3 and 5. Overall, the coefficients which can be understood as the longitudinal self-enhancement effects imply that the self-concept does not affect subsequent achievement in an opposite domain. Moreover, the RI/EM showed a nil effect from prior verbal self-concept to subsequent math self-concept between grade 3 and 5, and between math self-concept and subsequent verbal self-concept. The coefficients demonstrate that the self-concepts across the domain and time are not related, indicating that temporal comparison processes within one domain do not affect self-concept formation in the other domain. Similarly, the cross-sectional analyses also evidenced small correlational coefficients between verbal and math self-concept in grade 3.

Study II. The Developmental Relations of Students' Reading Self-concept and Reading Competence: Examining Reciprocal Relations and Ethnic-Background Patterns

The second study essentially investigated the reciprocal relations between reading self-concept and competence from grade 1 to 5 and addressed two specific core research queries. The first one examined whether there were reciprocal relations among students' reading self-concept and competence during primary school-age students and whether the reciprocal relations differ by ethnic/migration-backgrounds (White/Black/Hispanic/Asian) of students. The second question focused on whether there were ethnic differences in the students' reading self-concept and competence during primary school-age (grades 1 to 5), and how the magnitude of such differences compares to one another across time points. To deliver practical answers to the defined research questions, a longitudinal (ECLS-K), three waves, and two-year autoregressive cross-lagged paneled design were used in a large, representative sample of U.S. primary school children. The analysis was conducted without and with controlling for the highest SES of the parents at a given time point. Overall, the model fit indices were satisfactory for the longitudinal and multi-group MI models. I have documented all syntax online at the OSF (Center for Open Science, 2018; https://osf.io/8qsf9/?view_only=ff61d09d3dba49b5b92ae4b22d91e6ac), to make the present analyses transparent and reproducible (Nosek et al., 2015).

The reciprocal relations were revealed between reading self-concept and competence from the paths of grades 3 to 5 (for detail account please see appendix 2, result section); that is, early reading self-concept positively and significantly predicted the individual difference in later reading competence over and above the effect of prior reading competence (the self-enhancement part). Also, prior reading competence positively and significantly predicted the individual differences in later reading self-concept beyond the effect of prior reading self-concept at both time periods (the skill development part). A comparison of the reciprocal relations between reading competence and self-concept more strongly supported the skill development part within the REM, that is, larger effects of reading competence on self-concept were found than vice versa. A comparable pattern of results was also present when family SES was taken into account in the model.

In this set of analyses, the researcher executed multi-group REM analyses between reading self-concept and competence in grades 1 to 5 (please see appendix 2, result section). Generally, the pattern of the causal cross-path model did not show the reciprocal relations between reading self-concept and competence between grades 3 and 5 (except for White students). However, some significant and strong unidirectional path effects (i.e., skill development) were shown from reading competence to self-concept along all paths from grade 1 to 5 for all ethnic backgrounds. We also conducted further analyses to test significances for possible ethnic background differences in the pattern of reciprocal relations using an ethnic-invariant model. The result revealed that the reciprocal relations between reading self-concept and competence did vary across the ethnic-origins of students; however, note that the REM was observed for White students only. All of these results were also evident when the researcher considered SES of the family in the model. While longitudinally the descriptive results revealed comparable reading self-concept across the ethnic-backgrounds, there was a larger ethnic disparity for White and Asian over Black and Hispanic family students in reading competence.

Study III. Development of Sex Differences in Math Achievement, Self-Concept, and Interest From Grade 5 to 7

The third study fundamentally focused on examining the sex-related reciprocal effects (or covariance structure) among math achievements (i.e., standardized test-scores and teacher-assigned grades), self-concept, and interest from grade 5 to 7. More specifically, the study used cross-lagged panel analyses across sex groups to study the interplay of both self-concept and interest and achievement. In addition, the study extended the examination of sex differences beyond the bivariate relations to the latent mean structure of these constructs and described their trajectories. To test the aforementioned objectives, data from 2,342 German fifth- to seventh-grade students who participated in a large-scale longitudinal study BiKS (*Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter*; Educational processes, competence development, and selection decisions in preschool and school-age children, see Artelt, Blossfeld, Faust, Rossbach, & Weinert, 2013) was used. Before the main analyses, the longitudinal across time points (grade 5, 6, and 7) and multi-group MI were tested, and the model fit indices were satisfactory for both models. To make the present analyses transparent and reproducible (Nosek et al., 2015), we have provided all syntax online at the OSF (Center for Open Science, 2018, https://osf.io/aesbc/).

The cross-lagged panel analysis result revealed that math self-concept and both achievement measures are *reciprocally* related, that is, *mutually* reinforcing (for detail account please see appendix 3, result section). Besides, although the reciprocal effects were positive and significant for both achievement measures, they were more pronounced for grades than for standardized test-scores. The present study found a non-significant effect of interest on achievement if the self-concept was juxtaposed and not juxtaposed into the model in both achievement measures. Altogether, the present results provided further evidence for the mutual relationship between math self-concept and achievement (reciprocal effect) for both math achievement measures (i.e., test-scores and grades). The study also tested whether there are any sex-related differences in the reciprocal relations between variables (i.e., motivational constructs and school achievement). The study results provided strong support for the generalizability of the REM across sex groups. Differently put, sex is not a potential moderator of the reciprocal effects. Notwithstanding the similarities in the reciprocal relations of motivation and achievement across sex groups, there were substantial sex differences in the mean level of math self-concept and interest in a stereotyped way.

Discussion

The present dissertation aimed to provide further insight into the complex developmental relationship between academic self-concept and achievement. Specifically, the first goal of the dissertation was to examine the developmental interplay between self-concept and a standardized achievement measure within and across two domains (i.e., verbal and math) in primary school-aged students. The second goal was to investigate the reciprocal relations between verbal self-concept and competence and the moderating role of students' ethnic background from grade 1 to 5. The last contribution dealt with the development of sex differences in the reciprocal relations of academic self-concept and achievement in secondary school students using high-and low-stake achievement measures. Ultimately, the three studies comprised in this dissertation focused particularly on the development of academic achievement motivation in primary and secondary school-aged students and the role of relevant moderators in the achievement-motivation relationship. In the three studies, the main analyses were conducted after assuring longitudinal and group-specific measurement invariances (i.e., sex and ethnic-background). As in longitudinal studies, measurement invariance testing across time points is pivotal before making valid statements about the developmental aspect (Sewasew et al., 2018).

Developmental relationships between academic self-concept and achievement in primary school-aged students: The results of Study I offered strong empirical support for the reciprocal relations in both domains (i.e., verbal and math achievement motivation) during primary school education, that is, competence and the accompanying self-concept are mutually dependent over time. The mutual dependency does not imply that the effects are equal in size—the effects of academic competence on self-concept (i.e., skill development effect) were significantly stronger than the effects of academic self-concept on competence (i.e., self-enhancement effect), which corroborated the finding of Wolff et al. (2018); that is, positive paths were found from the change in math/language achievement to the corresponding self-concepts in primary school. The results also add to the existing evidence of the reciprocal relations between multiple academic self-concepts and achievement over time, combined with predominant effects of skill development effects in primary school-age students (Chapman & Tunmer, 1997; Guay et al., 2003; Helmke & Van Aken, G, 1995; Skaalvik & Valås, 1999).

Besides, the effects of the "classical" I/E frame of reference was also observed in the present study, thus, backing-up previous cross-sectional research with primary school students (Lohbeck & Möller, 2017; Pinxten et al., 2015). Moreover, in line with the RI/EM, achievement predicted subsequent academic self-concept in contrasting domains, which replicates previous findings with regard to primary school students (Wolff, et al., 2018). Comparing the magnitude of the cross-sectional with the longitudinal effects, it seems that the effects are not necessarily failing, which is important for considering the long-term effects of dimensional comparison. The only other RI/EM study in primary school (Wolff, et al., 2018) found negative, but non-significant effects leading from achievement levels and changes to non-corresponding self-concepts, which might be due to sampling size issues. This is important evidence, since most of the literature on the individual I/EM—both cross-sectional and longitudinal—focuses on secondary school students (Chen et al., 2013; Möller, Pohlmann, Köller, & Marsh, 2009; Möller et al., 2011, 2014; Niepel et al., 2014), omitting the important first years of schooling in self-concept formation (Eccles, Wigfield, Harold, & Blumenfeld, 1991; Marsh, 1985, 1989).

In the present investigation of primary school students, the study essentially found the same relations between self-concept and achievement, both cross-sectional and longitudinal, which indicates that the same cognitive mechanisms are in place. Differently put, already primary school students possess a basic understanding of their own abilities in terms of differentiation, that is, they clearly separate between academic domains and do not adhere to a monolithic academic self-concept. The results can also be taken as a cross-cultural validation on the RI/EM for the USA student sample because hitherto all available studies are from either Europe (Belgium, Germany) or Asia (Taiwan). The present study makes an important contribution to the existing literature by the virtue of expanding the number of replication studies. Replication studies are vital in science in general and in educational psychology in particular since the results not only have theoretical implications but also often inform practical educational consequences. Therefore, it is troublesome if only a small percentage of the publications deal with replications (see Makel, Plucker, & Hegarty, 2012).

The result of Study II showed the full reciprocal effect between reading self-concept and competence among primary-aged children in the overall model, even after partialling out family SES. That means the verbal self-concept and competence are developmentally interlocked—reciprocally related and mutually reinforcing over time. When the existing literature in the area is examined, this result is the first one to provide full support for REM in a primary education sample using comprehensive domain-specific standardized test scores, aside from Guay et al. (2003). However, our study is different from theirs in that we used a domain-specific self-concept and standardized test scores in a large representative U.S. sample with/without consideration of SES of the family. However, the study by Guay and colleagues was confined to domain-general achievement measures (i.e., the cumulative results of the teacher's rating in reading, writing, and math) and perceived competence beliefs of the small French student sample. Compared to other secondary school studies, so far only Retelsdorf et al. (2014) have found some reciprocal relations at the beginning of secondary education (i.e., between grades 5 and 6), but not along all paths of analyses. Similarly, consistent with our study, some studies of REM (e.g., Chen et al., 2013; Möller et al., 2014; Niepel et al., 2014) found reciprocal effects between a verbal grade and self-concept using teacher-assigned grades. This partial similarity of the results could be considered *conceptual replications* of our findings in primary (e.g., Guay et al., 2003) and secondary samples (Chen et al., 2013; Möller et al., 2014; Niepel et al., 2014), and evidence of cross-cultural and age comparisons. In this regard, Bonett (2012) stated, "replication evidence is the gold standard by which scientific claims are evaluated" (p. 410). Unfortunately, only 1% of the publications in psychology are replications (Makel et al., 2012).

Developmental relationships between academic self-concept and achievement in secondary school-aged students: The result of Study III evidenced that math self-concept and both achievement measures are *reciprocally* related, that is, *mutually* reinforcing. This result corroborates other empirical findings in the area (e.g., Ganley & Lubienski, 2016; Marsh et al., 1999, 2005; Marsh & Martin, 2011; Niepel et al., 2014; Pinxten et al., 2014). Although the reciprocal effects were positive and significant for both achievement measures, they were more pronounced for grades than for standardized test-scores. This result agrees with the notion that grades are more tangible and salient for students' self-beliefs and, therefore, influence self-concepts more strongly than standardized test scores (Marsh et al., 2005; Möller et al., 2014; Pinxten et al., 2010) which do not provide any feedback. The present study also revealed a non-significant effect of interest on achievement if the self-concept was juxtaposed and not juxtaposed into the model in both achievement measures. This result challenges the implicit assumption that math interest asserts an effect on achievement in math and vice versa. However, previous research using standardized test-scores (e.g., Ganley & Lubienski, 2016; Pinxten et al., 2014) also found non-reciprocal effects between math competence and interest when considered in a model with self-concept simultaneously (see also Marsh et al., 2005). One explanation for the absence of effects is the high construct and empirical overlap between interest and self-concept, which often leads to collinearity issues. Moreover, interest has been shown to indirectly influence school competence (Hübner et al., 2017).

The results provided further evidence for the mutual relationship between math self-concept and achievement (reciprocal effect) for both math achievement measures (i.e., test-scores and grades). This study also evidenced the stronger effects of math self-concept on achievement using standardized test-scores and teacher-assigned grades (*self-enhancement effects*). Following the guidelines outlined by Marsh and his colleagues (Marsh & Craven, 2006; Marsh et al., 2005), the present study supports the view that math self-concept is a significant and consistent predictor of math achievement. This is in contrast to the findings of Ganley and Lubienski (2016) and Pinxten et al. (2014), who reported causal relations (using standardized test-scores) in the opposite direction; in their results, that is, math competence is a consistent predictor of self-concept (*skill development effects*).

Overall, in the three studies, the causal predominance effects were evidenced, for example, *skill development* part (study I & II for primary school-aged children) and *self-enhancement* part (study III for secondary school-aged students). This could be attributed to the aged-related dynamic of academic self-concept development. On the one hand, in an early age, children's self-concept is characterized by exaggerated estimations of personal ability (Dweck, 2002) and academic self-concept may rely on prior achievement and not vice versa; the result is a skill development effect for younger children and a reciprocal effect for older age groups (Fraine, Damme, & Onghena, 2007). On the other hand, with the increase of age, young children's self-concept becomes more firmly formed (Marsh, 1989). Once self-concepts become more firmly established, the relation to achievement is likely to become reciprocal in that students with a higher academic self-concept set about a task with more confidence, and success is likely to strengthen their academic self-concept (Skaalvik & Hagtvet, 1990).

The moderational role of ethnic-background and sex in determining academic self-concept and achievement: In the present dissertation, beyond testing the longitudinal relationship between academic self-concept and achievement, the role-relevant moderators (i.e., ethnic background and sex of students) of achievement-motivations were examined. Study II tested whether there are any ethnic-related background differences in the reciprocal relations between variables (i.e., reading self-concept and competence). The study found strong support for the non-generalizability of REM across ethnic background groups; stated differently, ethnic background is a potential moderator of the reciprocal effects in primary school-aged children. The result is consistent with recent studies of reading competence and interest in the German secondary school sample (Miyamoto et al., 2017; Schaffner et al., 2016), with the reciprocal relations being evidenced only for native secondary school students. However, our finding is contrary to another analysis of Germany secondary school samples (Möller et al., 2014) which showed that the reciprocal relations between verbal achievement and self-concept are not restricted to whether students have an immigration background or not. The present study diverges from their study in various ways which may have led to dissimilar findings. For example, Möller et al.'s study (2014) involved secondary school students in Germany, whereas the present study focused on primary school children in the U.S. In the latter case, social stratification effects are more pronounced in academic performance because of ethnic-group disparities, and eventually affect achievement-motivation relationships. In addition, verbal competence was measured based on native language for the German sample, and most of those study participants were native students. For the U.S. sample, English is a second language for many as a significant number of children have immigrant family backgrounds. As a result, the academic disparity is more likely to occur because of the language barrier and integration challenges.

Also, in line with the aspiration-achievement paradox approach, the study found comparable reading self-concept regarding students' ethnic backgrounds across different grades; however, this is not true for reading competencies, as a large effect size of ethnic imbalance was evidenced particularly for White and Asian over Black and Hispanic background students. This result fits with most of the past research examining this question within the ethnic achievement-motivation gaps area (Bécares & Priest, 2015; Miyamoto et al., 2017; Ogbu, 1978; Salikutluk, 2016; Schaffner et al., 2016); all stressed the role of parental *aspirations or optimism* for the positive development of children's self-concept in reading. On the other hand, the overall poor performance of non-White students in reading competence, but not in reading self-concept, is partly explained by the *stereotype threat* theory—whereby the negative stereotype of a group influences their members' performance (Steele & Aronson, 1995). In this regard, Ogbu (1978) asserted that non-White students often perceive that efforts in school frequently do not have outcomes equivalent to those for members of socially-dominant groups, such as Whites. Accordingly, non-White students perceive the opportunity structure (e.g., success in education) differently from Whites and consequently tend to put less effort and commitment into their schoolwork, ultimately performing less well on the average than do majority students.

The third study further tested whether there are any sex-related differences in the reciprocal relations between variables (i.e., motivational constructs and school achievement). In contrast to the predictions derived from the *differential socialization hypothesis*, which assumes differential paths for the sexes (Marsh, 1989), the results provided strong support for the generalizability of the REM across sex groups. Shortly put, sex is not a possible moderator of the reciprocal relations. This result is along the lines of erstwhile meta-analytical investigations (Huang, 2011; Valentine, DuBois, et al., 2004) and recent studies of REM (see Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014) which confirmed no sex-associated change in the reciprocal effects. The absence of sex group differences in REM is partly attributed to the dichotomous nature of the construct; however, it is supposed that continuous grouping (e.g., interest) could result in differentiation between achievement-motivations.

Notwithstanding the similarities in the reciprocal relations of motivation and achievement across sex groups, there were substantial sex differences in the mean level of math self-concept and interest in a stereotyped way. Similarly, Marsh et al. (2005) reported substantially higher levels of math self-concept and interest for males. Moreover, in the current study, longitudinal sex differences remain stable for self-concept and seem to increase interest throughout secondary school. This result is in contrast to several studies showing that sex differences seem to decline in academic interest and self-concept (see Arens & Hasselhorn, 2014; Dotterer, McHale, & Crouter, 2009; Fredricks & Eccles, 2002; Jacobs et al., 2002; Pinxten et al., 2014, but see also Nagy et al., 2010, for stable sex differences), which might be due to the specific sample or the cultural context. Two things seem important to stress: First, sex-related mean differences in motivational constructs are crucial because interest and self-concept are more decisive than potential mean differences in achievement for educational career choices such as the enrollment in STEM subjects in university (Hübner et al., 2017; Marsh et al., 2005). Second, moderate and increasing differences in math self-concept and interest do not transfer to math achievement. Statistically speaking, this is sensible because the mean structure and the variance-covariance structure are independent and, thus, might independently affect students' competence and educational routes.

Implications and Limitations

The present dissertation has many significant theoretical and practical implications. The study provided strong support for positive developmental relations between academic achievement and self-concept within a domain and negative relations across two non-matching domains in a representative sample of primary-aged children with rigorous state-of-the-art analyses. Besides, to the best of the researcher's knowledge, the dissertation is one of the few studies to test reciprocal effects between verbal self-concept and achievement (i.e., standardized test-scores) without and with controlling SES in the subdomain of reading in a sample of primary school-aged students. Moreover, the study has provided evidence for the mutual relationship between math self-concept and achievement (reciprocal effect) for both high and low-stake math achievement measures (i.e., grades and test-scores) in secondary school students. Furthermore, the causal predominance of the self-*enhancement effects*—stronger effects of math self-concept on achievement using standardized test-scores and teacher-assigned grades in secondary school-aged students—have not previously been evidenced in the literature (see Marsh et al., 2005 for an exception).

Academic self-concept formation is not only important for the actual performance of students, for the comparisons they make are also essential antecedents of academic self-concept. In light of the RI/EM presumed model, the present dissertation offered support for the underlining mechanisms of academic self-concept formation and academic achievement evaluation involving three types of cognitive processes: social, temporal and dimensional comparisons. These three comparison processes influence the formation of self-concept; to some degree, they work independently (solely) and to some degree they are interdependent (interplay, Möller & Marsh, 2013). At this point, it is worth mentioning that these cognitive processes of comparisons are rather interpretive of academic self-concept formation and achievement development evaluation. Nonetheless, they are plausible assumptions of how social, dimensional and temporal achievement comparisons occur, and explain the underlying mechanisms in academic self-concept formation, at least partly without causal assertions (Möller & Marsh, 2013; Wolff, Nagy, et al., 2018).

Two divergent results are evidenced by group differences of achievement-motivation relations, on the one hand, and the lack of sex-related differences in the reciprocal relations between academic self-concept and achievement (i.e., math) in secondary school-aged students. This is quite an interesting finding since it contradicts theoretical assumptions about sex-related differential pathways, and moreover, is in stark contrast to substantial differences in the mean structure across sex groups. Nonetheless, it implied that sex-based intervention schemes, particularly those designed for girls should be implemented to enhance their motivational abilities. On the other hand, the study has revealed the academic disparity based on ethnic-background in verbal competence. The social stratification effects due to poverty should not be an excuse for poor verbal competence, as many students at the wealthier schools and from educated families were not performing very well either. Instead, if students are taught the right way (i.e., learning to read) their competence would be improved as well (Hanford, 2018). In the conventional curriculum, the traditional philosophy (i.e., reading to learn) of balanced literacy is to provide children a lot of good books, and through little direction and sufficient exercise, they will become good readers. However, this does not appear to be sustainable enough to improve performance in reading for many children. Hence, teachers need to be taught the science of learning to read (i.e., how to connect sounds with letters—phonics) and avoid their deeply held beliefs about reading (i.e., reading is a natural development, similar to learning to talk). Simultaneously, explicit and systematic phonics instruction is very crucial for students, to teach them about learning to read proficiently (Hanford, 2018).

Taken together, the study suggests that academic achievement and self-concept should be fostered simultaneously in primary and secondary school-age students (see also Marsh et al., 2005). However, there is a need to caution against an “either-or” approach—either fostering prior achievement, which affects subsequent academic self-concept (skill development) or fostering prior academic self-concept, which affects subsequent achievement (self-enhancement). For instance, if parents or teachers improve students’ academic self-concepts without cultivating their achievement, the student improvements in academic self-concept will not be long-lasting. This fact is particularly relevant for primary school children because boosting academic self-concept and skills is most influential at this developmental stage. After primary school, the link between self-concept and achievement most likely weakens (Retelsdorf et al., 2014) due to the fact that the academic self-concepts become more stable and less influenced by prior achievement (Chen et al., 2013), and negative effect of school average achievement on academic self-concept, as students who attend high-ability classes and schools tend to have lower academic self-concepts than do equally able students who attend mixed or low ability classes and schools (Marsh et al., 2017). Increased peer pressure also play a role in secondary education. As a result, parental support, academic teaching, and educational programs, in general, should nourish and foster the academic self-concepts and academic skills jointly in the early years of primary schooling (Retelsdorf et al., 2014).

Furthermore, the extended RI/EM results show negative and significant effects of prior academic achievements (e.g., verbal) on contrasting subsequent academic self-concepts (e.g., math), confirming the interplay of dimensional and temporal processes. If teachers and parents are aware of students' internal and temporal comparisons, the negative effects of academic achievement on the non-matching self-concept can perhaps be mitigated. Therefore, it is important to be aware of and reflect on the devaluating potential of good grades in a specific subject on students' skills in a non-matching subject (Chen et al., 2013; Marsh & Köller, 2004).

Notwithstanding the above pertinent implications of the study, a number of limitations of this dissertation should be noted. First, it is noteworthy that the term "causality" is normally applied in self-concept investigations in which causal effects are assumed to refer to the reciprocating relations between academic self-concepts and achievements, but do not directly entail causality (see also Marsh et al., 2017). Nonetheless, as a side note, the results of interventions targeting girls' self-concept in STEM subjects (Dresel & Ziegler, 2006; Ziegler & Heller, 2000) are promising and provide strong evidence for the major role of self-concept and interest in knowledge acquisition and for educational career choices (Sewasew et al., 2018).

Second, social desirability might affect the results, as the usage of self-report measures of academic self-concept might reflect self-promotion. Self-promotion might happen if study participants tend to overrate their positive traits and underrate their negative traits (Stangor, 2011). Besides, the item wording can considerably change the results (Sewasew & Schroeders, 2019), for example, the item "I think I'm good in math" could be either "I think I'm good in math in comparison to my classmates" if the focus is on social comparison processes, or "I think I'm good in math compared to my previous achievements" if the focus is on the developmental aspect. Third, hitherto studies have reported a range of factors (beyond ethnic-origin and sex) that may influence the relationship between academic self-concept and achievement, for example, educational placement settings and class achievement level (Marsh et al., 2017). This study did not exclude the potential effects of these factors on the relationship between academic self-concept and achievement.

Finally, it is of note that this dissertation used sample students from two countries (the U.S. and Germany). As a result, the researcher was not able to make an age-matched parallel cross-cultural comparison, as the data are from primary and secondary school-aged students, respectively. However, given that the studies are both longitudinal; overall the results of this dissertation give a hint that academic self-concept and achievement are positively and developmentally related across culture, even though somehow there are cultural similarities in the two sample study settings. Nonetheless, a growing number of studies have found support for the reciprocated within- and across-domain relations between achievement-motivations in different countries and cultural groups (see also Guo et al., 2018; Marsh & Hau, 2004).

Final Thoughts

Overall, the present dissertation amplifies the substantial role of academic self-concept for student success, academic achievement, and as a factor that has desirable effects as a precursor as well as a consequence (i.e., it works as a moderator of essential outcomes). Essentially, it is greatly desirable for students to have a strong and balanced academic self-concept and achievement for a successful school career. However, in reality, that is not the case for many students, as several studies verified that there is an imbalanced relationship between these constructs. Nonetheless, ample studies report redundant results of the effects of academic achievement on self-concept (*skill development* effect); any less sweeping reciprocal relationship results due to small effect size are still meaningful, as they give a hint of the role of academic self-concept for students' success. In fact, it is clear that having academic ability, for example being clever, inspires self-confidence, although this may only apply while the success is easy and continuous. The decisive issue in academic success is how students perceive their academic ability over the course of their school career, particularly how they deal with academic failure and maintain positive attitudes about their academic ability. In the present context, all of the three papers confirmed the presence of reciprocal effects in primary and secondary school-aged students (i.e., at least give a clue for the importance of academic self-concept on achievement); particularly the third paper magnified the strong causal predominance of academic self-concept (*self-enhancement* effect) overachievement in low and high-stack achievement measures. That means the role of motivational ability is highly significant for academic success. Consequently, teachers and parents are strongly encouraged to work on building strong academic self-concept and to think through academic self-concept inferences when reaching academic decisions.

In addition, this dissertation cross-sectionally and longitudinally extends the *internal comparison* processes, resulting in negative cross-domain effects on academic self-concept, and *external comparison* processes, resulting in positive within-domain effects on academic self-concept. Therefore, there is a pattern of mutual effects that strengthen the *domain specificity* of these constructs and the multidimensionality of academic self-concepts. Hence, teachers and parents should not assume that academic self-concepts are corresponding across domains, and should constantly gauge the level to which their assessment of a student's competence in a certain academic subject is correct.

Moreover, the moderators of academic achievement-motivations have shown differences between students' academic performance (e.g., verbal) because of background disparities in primary school years; however, relatively comparable academic self-concept has been demonstrated in primary and secondary school years. Thus, the study stresses the importance of attaining ethnic achievement parity through eliminating social stratification effects and including intervention based schemes (e.g., learning to read rather than reading to learn) to avoid achievement deficits at early stages of schooling. In practical terms, the study underlines the need to use the leverage of positive motivational ability to translate into better academic competence. For example, applying deeper learning (via collaborative problem-solving) approach to enhance and develop a strong language self-concept, in turn, may promote students' academic success.

As a final point, the following themes would be avenues recommended for future research: first, a broader and deeper understanding of the formation of academic self-concept is still at stake, especially to thoroughly elucidate the interplay between academic self-concept and academic achievement across domains, time, and school/class levels in primary and secondary school-aged children (i.e., the critical period in students' cognitive development, academic self-concept formation and academic skill mastery). Second, a few previous longitudinal studies on the social and dimensional comparisons have been done; the present study drew on longitudinal data. Consequently, the strength of the study findings might be corroborated by well-designed quasi-or true experimental studies to better comprehend the underlying causal mechanisms between academic self-concept and achievement. Lastly, even though sex disparities in math competence have dropped in the past few years, sex differences in involvement in math subjects and professions relating to math still remain. Especially, the presence of small sex differences should not be ignored, as such achievement gaps may lead to larger differences later in the educational career of students. Equally, an overemphasis of the conditional mean analyses may have overshadowed the opportunity to detect when and how the sex differences in academic achievement emerge and may result in proposing overly-simplistic interventions. Hence, it is imperative to determine the genesis of sex differences in the STEM subjects at different proficiency levels of the achievement distribution so as to address the sex-based imbalance in STEM jobs.

Funding Statement

This work was supported by the Bamberg Graduate School of Social Sciences which is funded by the German Research Foundation (DFG) under the German Excellence Imitative (GSC1024).

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Arens, A. K., & Hasselhorn, M. (2014). Age and gender differences in the relation between self-concept facets and self-esteem. *The Journal of Early Adolescence*, *34*, 760–791. doi: 10.1177/0272431613503216.

Artelt, C., Blossfeld, H.P., Faust, G., Rossbach, H.G., & Weinert, S. (2013). *Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor-und Grundschulalter* (BiKS-8-14) [Educational processes, competence development and selection decisions in preschool and school age (BiKS-8-14)]. Version: 2. IQB - Institute for Quality Development in Education (Dataset). http://doi.org/10.5159/IQB_BIKS_8_14_v2

Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among eighth-grade students: Findings from an Intersectionality Approach. *PLOS ONE*, *10*(10), e0141363. https://doi.org/10.1371/journal.pone.0141363

Bem, D. J. (1972). Self-Perception Theory11: Development of self-perception theory was supported primarily by a grant from the National Science Foundation (GS 1452) awarded to the author during his tenure at Carnegie-Mellon University. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1–62). https://doi.org/10.1016/S0065-2601(08)60024-6

Bonett, D. G. (2012). Replication-Extension Studies. *Current Directions in Psychological Science*, *21*(6), 409–412. https://doi.org/10.1177/0963721412459512

Brunner, M., Gogol, K. M., Sonnleitner, P., Keller, U., Krauss, S., & Preckel, F. (2013). Gender differences in the mean level, variability, and profile shape of student achievement: Results from 41 countries. *Intelligence*, *41*(5), 378–395. https://doi.org/10.1016/j.intell.2013.05.009

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*(2), 136–145. https://doi.org/10.1037/0022-0663.69.2.136

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *The British Journal of Educational Psychology*, *67 (Pt 3)*, 279–291; discussion 339-343.

Chapman, James W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology*, *67*(3), 279–291. https://doi.org/10.1111/j.2044-8279.1997.tb01244.x

Chen, S.-K., Yeh, Y.C., Hwang, F.M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. *Learning and Individual Differences*, *23*, 172–178. https://doi.org/10.1016/j.lindif.2012.07.021

Craven, R. G., & Marsh, H. W. (2008). The centrality of the self-concept construct for psychological wellbeing and unlocking human potential: Implications for child and educational psychologists. *Educational and Child Psychology*, *25*(2), 104–118.

Dotterer, A. M., McHale, S. M., & Crouter, A. C. (2009). The development and correlates of academic interests from childhood through adolescence. *Journal of Educational Psychology*, *101*(2), 509–519. https://doi.org/10.1037/a0013987

Dresel, M., & Ziegler, A. (2006). Langfristige Förderung von Fähigkeitsselbstkonzept und impliziter Fähigkeitstheorie durch computerbasiertes attributionales Feedback. *Zeitschrift Für Pädagogische Psychologie*, *20*(1/2), 49–63. https://doi.org/10.1024/1010-0652.20.12.49

Dweck, C. S. (2002). The development of ability conceptions. *Development of Achievement Motivation* (pp. 57–88). https://doi.org/10.1016/B978-012750053-9/50005-X

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1991). Age and gender differences in children's self- and task perceptions during elementary school, *Child Development*, *64* (3), 830 - 847.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fraine, B. D., Damme, J. V., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, *32*(1), 132–150. https://doi.org/10.1016/j.cedpsych.2006.10.005

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*(4), 519–533. https://doi.org/10.1037/0012-1649.38.4.519

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics —A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics.

European Journal of Psychology of Education, *22*(4), 497. https://doi.org/10.1007/BF03173468

Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, *47*, 182–193. https://doi.org/10.1016/j.lindif.2016.01.002

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

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

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Hanford, E. (2018). Why aren’t kids being taught to read? Retrieved September 15, 2018, from https://www.apmreports.org/story/2018/09/10/hard-words-why-american-kids-arent-being-taught-to-read

Harter, S. (2012). *The construction of the self: A developmental perspective* (First edition). New York: The Guilford Press.

Helmke, A., & Van Aken, G, A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, *87*(4), 624–637. https://doi.org/10.1037/0022-0663.87.4.624

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Hübner, N., Wille, E., Cambria, J., & Trautwein, U. (2017). Maximizing gender equality in STEM by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM. *Journal of Educational Psychology,*109, 993-1009. https://doi.org/10.1037 / edu0000183.

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*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. *Journal of Educational Psychology*, *107*(4), 1086–1101. https://doi.org/10.1037/edu0000021

Kao, G., & Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth. *Social Science Quarterly*, *76*(1), 1–19.

Kigel, R. M., McElvany, N., & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation: A longitudinal study. *Learning and Instruction*, *35*, 73–84. https://doi.org/10.1016/j.learninstruc.2014.10.001

Lohbeck, A., & Möller, J. (2017). Social and dimensional comparison effects on math and reading self-concepts of elementary school children. *Learning and Individual Differences*, *54*, 73–81. https://doi.org/10.1016/j.lindif.2017.01.013

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542. https://doi.org/10.1177/1745691612460688

Marsh, H. W. (1985). Age and sex effects in multiple dimensions of preadolescent self-concept: A replication and extension. *Australian Journal of Psychology*, *37*(2), 197–204. https://doi.org/10.1080/00049538508256398

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1989a). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi.org/10.1037/0022-0663.81.3.417

Marsh, H. W. (1989b). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi.org/10.1037/0022-0663.81.3.417

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., … Abu-Hilal, M. M. (2015). The Internal/External Frame of Reference Model of Self-Concept and Achievement Relations: Age-Cohort and Cross-Cultural Differences. *American Educational Research Journal*, *52*(1), 168–202. https://doi.org/10.3102/0002831214549453

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and... *Educational Psychologist*, *34*(3), 155–167. https://doi.org/10.1207/s15326985ep3403_2

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. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of east and West German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology*, *29*(3), 264–282. https://doi.org/10.1016/S0361-476X(03)00034-1

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Pekrun, R., Murayama, K., Arens, K. A., Parker, P. D., Guo, J., & Dicke, T. (2017). An integrated model of academic self-concept development: academic self-concept, grades, test scores, and tracking over six years. *Developmental Psychology*. Retrieved from http://centaur.reading.ac.uk/71299/

Marsh, H. W., & Redmayne, R. S. (1994). A multidimensional physical self-concept and its relations to multiple components of physical fitness. *Journal of Sport and Exercise Psychology*, *16*(1), 43–55. https://doi.org/10.1123/jsep.16.1.43

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*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Mickelson, R. A. (1990). The attitude-achievement paradox among black adolescents. *Sociology of Education*, *63*(1), 44–61.

Miyamoto, A., Pfost, M., & Artelt, C. (2017). Reciprocal relations between intrinsic reading motivation and reading competence: A comparison between native and immigrant students in Germany: Reading Motivation and Reading Competence. *Journal of Research in Reading*. https://doi.org/10.1111/1467-9817.12113

Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, *93*(4), 826.

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research*, *79*(3), 1129–1167. https://doi.org/10.3102/0034654309337522

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 0002831211419649. https://doi.org/10.3102/0002831211419649

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education*, *21*(4), 467–487.

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, *84*(4), 591–611. https://doi.org/10.1111/bjep.12047

Müller-Kalthoff, H., Jansen, M., Schiefer, I. M., Helm, F., Nagy, N., & Möller, J. (2017). A double-edged sword? On the benefit, detriment, and net effect of dimensional comparison on self-concept. *Journal of Educational Psychology*, *109*(7), 1029–1047. https://doi.org/10.1037/edu0000171

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*(4), 1170–1191. https://doi.org/10.1037/a0036307

Ogbu, J. U. (1978). *Minority Education and Caste: The American System in Cross-Cultural Perspective*. Academic Press Inc.

Ogbu, J. U. (1993). Differences in cultural frame of reference. *International Journal of Behavioral Development*, *16*(3), 483–506. https://doi.org/10.1177/016502549301600307

O'Mara, A. J., 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*(3), 181–206. https://doi.org/10.1207/s15326985ep4103_4

Pastorin, E. ., & Doyle-Portillo, S. (2013). *What Is Psychology?: Essentials, Second Edition* (2nd edition). Cengage Learning.

Peterson, S. E., & Miller, J. A. (2004). Comparing the quality of students' experiences during cooperative learning and large-group instruction. *The Journal of Educational Research*, *97*(3), 123–134. https://doi.org/10.3200/JOER.97.3.123-134

Pinxten, M., de Fraine, B., van Damme, J., & D'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. *British Journal of Educational Psychology*, *80*(4), 689–709. https://doi.org/10.1348/000709910X493071

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, *84*(1), 152–174. https://doi.org/10.1111/bjep.12028

Pinxten, M., Wouters, S., Preckel, F., Niepel, C., De Fraine, B., & Verschueren, K. (2015). The formation of academic self-concept in elementary education: A unifying model for external and internal comparisons. *Contemporary Educational Psychology*, *41*, 124–132. https://doi.org/10.1016/j.cedpsych.2014.12.003

Pohlmann, B., & Möller, J. (2009). On the benefit of dimensional comparisons. *Journal of Educational Psychology*, *101*(1), 248–258. https://doi.org/10.1037/a0013151

Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction*, *29*, 21–30. https://doi.org/10.1016/j.learninstruc.2013.07.004

Rosenberg, M. (1965). Society and the Adolescent Self-Image. Morris Rosenberg. Princeton University Press, Princeton, N.J., 1965. xii + 326 pp. $6.50. *Science*, *148*(3671), 804–804. https://doi.org/10.1126/science.148.3671.804

Salikutluk, Z. (2016). Why Do Immigrant Students Aim High? Explaining the Aspiration–Achievement Paradox of Immigrants in Germany. *European Sociological Review*, *32*(5), 581–592. https://doi.org/10.1093/esr/jcw004

Schaffner, E., Philipp, M., & Schiefele, U. (2016). Reciprocal effects between intrinsic reading motivation and reading competence? A cross-lagged panel model for academic track and nonacademic track students: RECIPROCAL EFFECTS. *Journal of Research in Reading*, *39*(1), 19–36. https://doi.org/10.1111/1467-9817.12027

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of Reading Motivation and Their Relation to Reading Behavior and Competence. *Reading Research Quarterly*, *47*(4), 427–463. https://doi.org/10.1002/RRQ.030

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*(2), 390–433. https://doi.org/10.3102/0002831209350493

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Sewasew, D., Schroeders, U., Schiefer, I. M., Weirich, S., & Artelt, C. (2018). Development of sex differences in math achievement, self-concept, and interest from grade 5 to 7. *Contemporary Educational Psychology*, *54*, 55–65.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology*, *58*(2), 292–307. https://doi.org/10.1037/0022-3514.58.2.292

Skaalvik, E. M., & Valås, H. (1999a. Relations Among Achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, *67*(2), 135–149. https://doi.org/10.1080/00220979909598349

Stangor, C. (2011). *Research methods for the behavioral sciences*. Australia; Belmont, CA: Wadsworth Cengage Learning.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

Steele, Claude M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology* (Vol. 34, pp. 379–440). Elsevier.

Valentine, J. C., & DuBois, D. L. (2005). Effects of self-beliefs on academic achievement and vice versa. *International Advances in Self Research*, *2*, 53–77.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Villiger, C., Wandeler, C., & Niggli, A. (2014). Explaining differences in reading motivation between immigrant and native students: The role of parental involvement. *International Journal of Educational Research*, *64*, 12–25. https://doi.org/10.1016/j.ijer.2013.10.004

von Maurice, J., Dörfler, T., & Artelt, C. (2014). The relation between interests and grades: Path analyses in primary school age. *International Journal of Educational Research*, *64*, 1–11. https://doi.org/10.1016/j.ijer.2013.09.011

Weinert, F. E. (1999). *Concepts of competence*. OECD: Max Planck Institute for psychological research, Munich, Germany.

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology*, *110*(7), 1005–1025. https://doi.org/10.1037/edu0000248

Wolff, F., Nagy, N., Helm, F., & Möller, J. (2018). Testing the internal/external frame of reference model of academic achievement and academic self-concept with open self-concept reports. *Learning and Instruction*, *55*, 58–66. https://doi.org/10.1016/j.learninstruc.2017.09.006

Yoon, K. S, Eccles, J. S, & Wigfield, A. (1996). Self-concept of ability, value, and academic achievement: A test of causal relations. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1996. Washington: American Educational Research Association.

Ziegler, A., & Heller, K. A. (2000). Effects of an Attribution Retraining with Female Students Gifted in Physics. *Journal for the Education of the Gifted*, *23*(2), 217–243.

Appendix

List of Original Contributions

1. Sewasew, D., Schroeders, U. (2019). The developmental interplay of academic self-concept and achievement within and across domains among primary school students. *Contemporary Educational Psychology*, https://doi.org/10.1016/j.cedpsych.2019.03.009
2. Sewasew, D., Koester, L.S. (2019). The developmental dynamics of students' reading self-concept and reading competence: examining reciprocal relations and ethnic-background patterns. *Learning and Individual Differences, 79,* 102-111. https://doi.org/10.1016/j.lindif.2019.05.010
3. Sewasew, D., Schroeders, U., Schiefer, I., Weirich, S., Artelt, C. (2018). Development of sex differences in math achievement, self-concept, and interest from grade 5 to 7. *Contemporary Educational Psychology*, *54*, 55–65, https://doi.org/10.1016/j.cedpsych.2018.05.003.

The Developmental Interplay of Academic Self-Concept and Achievement Within and Across Domains Among Primary School Students

Daniel Sewasew[a]

Ulrich Schroeders[b]

[a] *Bamberg Graduate School of Social Sciences (BAGSS), University of Bamberg, Germany*

[b] *Institute of Psychology, University of Kassel, Germany*

The data used for this study were obtained from the "Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999" (ECLS-K), a project aimed at assessing educational, physical, and social development among children over time (Tourangeau, Nord, Le, Sorongon, & Najarian, 2009). All ECLS-K, kindergarten through eighth grade public-use files are available from the *National Center for Educational Statistics* (https://nces.ed.gov/ecls/dataproducts.asp).

Correspondence concerning this article should be addressed to Daniel Sewasew, Bamberg Graduate School of Social Sciences, University of Bamberg, Feldkirchenstraße 21, 96045 Bamberg, Germany, email: sewasewdaniel@gmail.com, phone: +49 (0)951 863 3617.

Abstract

The reciprocal internal/external frame of reference model (RI/EM) extends the internal/external frame of reference model (I/EM) over time and the reciprocal effects model (REM) across domains. The RI/EM postulates positive developmental relations between academic achievement and self-concept within a domain and negative relations across two non-matching domains (e.g., math and English). However, until now, empirical investigations of the RI/EM had only focused on secondary school students from specific countries. In the present study, we test whether the RI/EM also applies to primary school students and to students in the United States, by using a representative longitudinal data set: the Early Childhood Longitudinal Study-Kindergarten (ECLS-K: 1998-1999). We found positive reciprocal relations between academic self-concept and standardized test scores within a domain, whereas the effect of prior achievement on self-concept was much stronger (skill-development part) than the effect of self-concept on achievement (self-enhancement). Furthermore, we found negative effects of achievement on subsequent self-concepts across domains (I/E frame of references). Overall, the findings of the study strongly support the RI/EM for primary school students. Our results are compared to previous findings in the literature for secondary school students and are discussed with regard to self-concept formation in primary school.

Keywords: academic self-concept, academic achievement, reciprocal internal/external model, primary school students

The Developmental Interplay of Academic Self-Concept and Achievement Within and Across Domains Among Primary School Students

1. Introduction

Academic self-concept is one of the key constructs in the research on scholastic motivation and achievement (Marsh et al., 2017; Weinert, 1999). Academic self-concept is defined as the self-perception of a student's ability in an academic domain. It is a pivotal predictor of several educational outcomes, including individual future visions and motivation to learn (Guay, Marsh, & Boivin, 2003; Marsh & Craven, 2006), educational and occupational aspirations (Marsh, 1990), course selection (Marsh & Yeung, 1997) and career intentions and choices (Guay, Larose, & Boivin, 2004; Jansen, Scherer, & Schroeders, 2015). Thus, academic self-concept is closely intertwined with academic achievement and develops through temporal comparison (individual performance at an earlier point in time), dimensional comparison (performance across different subjects), and social comparison (standing within a class).

Two established models try to explain the relationship between self-concept and achievement more thoroughly: the *reciprocal effects model* (REM) and the *internal/external frame of reference model* (I/EM). The REM (Marsh & Craven, 2006) states that academic achievement and self-concept are mutually related over time, whereas the I/EM (Möller, Pohlmann, Köller, & Marsh, 2009) proposes that academic achievement positively influences a student's academic self-concept within a given domain (*external reference*), but negatively affects his or her self-concept in a non-corresponding domain (*internal reference*) at a given point in time. To simultaneously study the influence of all sources, both models have been combined into the reciprocal *internal/external frame of reference model* (RI/EM). This model considers the longitudinal interplay between academic self-concept and achievement across multiple domains (Möller, Retelsdorf, Köller, & Marsh, 2011).

The RI/EM incorporates all three types of comparison essential for building, maintaining, and shaping an academic self-concept: social, temporal, and dimensional comparisons. That is, students develop and sustain their self-concept by comparing their own achievement to one of their classmates (social comparison, Festinger, 1954), to their prior achievement (temporal comparison, Albert, 1977), and to their achievement in other subjects (dimensional comparison, Möller & Marsh, 2013). The classical REM covers the temporal and social comparisons, while the I/EM reflects social and dimensional comparisons. The RI/EM, therefore, extends the REM to include dimensional comparisons, while the I/EM is expanded to include temporal comparisons.

The empirical evidence for the RI/EM is still sparse, since it has only been developed recently (Chen, Yeh, Hwang, & Lin, 2013; Möller et al., 2011; Möller, Zimmermann, & Köller, 2014; Weidinger, Steinmayr, & Spinath, 2019; Wolff, Helm, Zimmermann, Nagy, & Möller, 2018). Moreover, this model has been studied predominantly in European samples (e.g., Belgium and Germany). However, this fact might not affect the generalizability of the model, at least for Western countries, given the huge body of literature demonstrating the cross-cultural applicability of its parts (for the I/EM, see Marsh et al., 2015; for the REM, see Marsh, Hau, & Kong, 2002). Of greater concern is that most of the existing studies only sampled secondary school-aged students, even though primary school is the most important period in forming achievement motivation (Guo, Marsh, Parker, & Dicke, 2018), given the fast progress of cognitive and academic skills in this period of time (Harter, 2012). In general, academic self-concept is subject to age-related dynamics and declines from a young age through adolescence (Marsh, 1989).

In primary school, children tend to overestimate their academic abilities, and their evaluation is often weakly correlated with external criteria such as grades and test scores (Marsh, 1985). In the course of primary education, students’ self-concept begins to differentiate and to rely more strongly on social comparison driven by feedback from teachers and parents (Marsh, 1989; Wigfield et al., 1997). Due to these fundamental changes in self-perception skills (Marsh, 1985), the increased importance of grades (Marsh, 1989), and the general progress in cognitive development during primary school (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002), it is important to examine to what extent the predictions of the RI/EM also apply to primary school students. Thus, the main purpose of the present study is to rigorously analyze the RI/EM in a representative sample of primary school-aged students in a different educational setting (i.e., the United States). Below, we will present a brief overview of the theories and research on academic self-concept and achievement in light of the REM, the I/EM, and their combination in the RI/EM, followed by our research questions in detail.

1.1. The Reciprocal Effects Model (REM)

Academic achievement and self-concept are strongly associated. However, the question of whether motivational or achievement variables are more important in the chronological sequence remains unanswered and is one with significant theoretical and practical consequences (see Marsh et al., 2017). Classical explanations of this question have centered around an either-or-view: either early achievement influences later academic self-concept via the social comparison process (*the skill development model*) or prior academic self-concept influences subsequent achievement directly (*the self-enhancement model*). A way of reconciling these two opposing conceptualizations is the REM, which states that early academic achievement affects later academic self-concept; in the same vein, prior academic self-concept affects subsequent academic achievement (Marsh & Craven, 2006).

There is a large body of literature demonstrating the usefulness of the REM for secondary school students (Chen et al., 2013; Möller et al., 2011; Möller et al., 2014; Niepel et al., 2014; Sewasew, Schroeders, Schiefer, Weirich, & Artelt, 2018). However, only a few studies have provided empirical evidence for the REM in a primary school setting. For example, Guay et al. (2003) found support for more or less constant mutual effects between achievements (i.e., reading, writing, and math) and competence beliefs (i.e., self-concept) over different age cohorts in their multi-cohort, multi-occasion study of 385 French Canadian primary school children. However, results opposing the REM are also documented in the literature: for example, only unidirectional effects for the skill development paths were present in a Norwegian primary and middle school sample for both math and verbal subjects (Skaalvik & Valås, 1999), and no reciprocal relations were obtained between verbal competence and self-concept for very young children between grades 1 and 2 (Chapman & Tunmer, 1997). Likewise, Helmke and Van Aken (1995) reported a unidirectional relation between math competence and self-concept in a primary school sample (grades 2 and 3) when using test scores and school grades. Moreover, a recent longitudinal study using the same US sample as in the present study has documented a consistent unidirectional effect of math achievement on math self-concept from 3rd, 5th, and 8th grade (Ganley & Lubienski, 2016). The lack of REM effects in primary school can partly be attributed to the prematurity of children's self-concept, as academic self-concept becomes more firmly established and stable with age (Chen et al., 2013).

1.2. The Internal/External Frame of Reference Model (I/EM)

The I/EM states that academic self-concept in a certain school subject is shaped by two evaluations: social comparison (or *external reference*), in which students relate their achievement to the achievement of other students within a certain school subject, and dimensional comparison (or *internal reference*), in which students evaluate their self-concept in a given subject based on their achievement in a non-matching domain (most often studied with math and language domains). Despite the fact that academic achievement in math and verbal domains is strongly associated, the respective self-concepts are only weakly correlated (Marsh, Gerlach, Trautwein, Lüdtke, & Brettschneider, 2007). In a comprehensive meta-analysis, Möller et al. (2009) integrated the results of 68 data sets with more than 125,000 participants, providing strong evidence for the I/EM. They reported a high positive correlation between math and verbal achievement (r = .67), while the correlation between the self-concepts was close to zero (r = .10). In line with an external frame of reference, the paths from math achievement to math self-concept (β = .61) and from verbal achievement to verbal self-concept (β = .49) were positive. Furthermore, in agreement with the internal frame of reference, the paths from math achievement to verbal self-concept (β = -.21) and from verbal achievement to math self-concept (β = -.27) were significantly negative. Besides the meta-analytical evidence, the I/EM is supported by many different methodological approaches: experimental studies (Möller & Köller, 2001; Müller-Kalthoff et al., 2017; Pohlmann & Möller, 2009; Wolff, Helm, et al., 2018), cross-sectional studies (Lohbeck & Möller, 2017; Marsh & Hau, 2004; Pinxten et al., 2015), and longitudinal field studies (Chen et al., 2013; Möller et al., 2011; Möller, Zimmermann, & Köller, 2014; Niepel et al., 2014; Wolff et al., 2018). Moreover, the I/EM has been extended to domains other than math and verbal(e.g., in sciences to biology, physics, and chemistry, Jansen, Schroeders, Lüdtke, & Marsh, 2015; Möller, Streblow, Pohlmann, & Köller, 2006).

1.3. The Integration of the REM and the I/EM into the RI/EM

The RI/EM combines dimensional comparison (i.e., students comparing their performance across two different domains, usually math and verbal), temporal comparison (i.e., students relating their current performance to prior performance in the same domain), and social comparison (i.e., students evaluating their performance in comparison to others). Merging the I/EM and the REM complements the comparisons of each individual model and gives a full picture of the underlying processes (Marsh & Köller, 2004): the REM lacks the cross-domain perspective, while the I/EM disregards the developmental aspect and self-enhancement effects. In Figure 1, the assumptions of the RI/EM are summarized: (a) *skill development effects* (β_{SD}, temporal and social comparison), (b) *self-enhancement effects* (β_{SE}, temporal and social comparison), (c) *external frame of reference effects* (β_{EF}, social comparison), (d) *internal frame of reference effects*, both cross-sectional (β_{IF}, dimensional comparisons) and longitudinal (β_{IFL}, temporal and dimensional comparisons), (e) *autoregressive effects*, (f) *reciprocal cross-domain effects of achievement*, (g) *cross-domain contrast effects of academic self-concepts*, and (h) *cross-domain contrast effects of academic self-concept on achievement*. As a side note, the skill development effects can also be described in terms of the I/EM, that is, as the result of social comparison (i.e., longitudinal external frame of reference effects), but in the longitudinal run (see Möller et al., 2011, 2014; Niepel et al., 2014; Wolff et al., 2018).

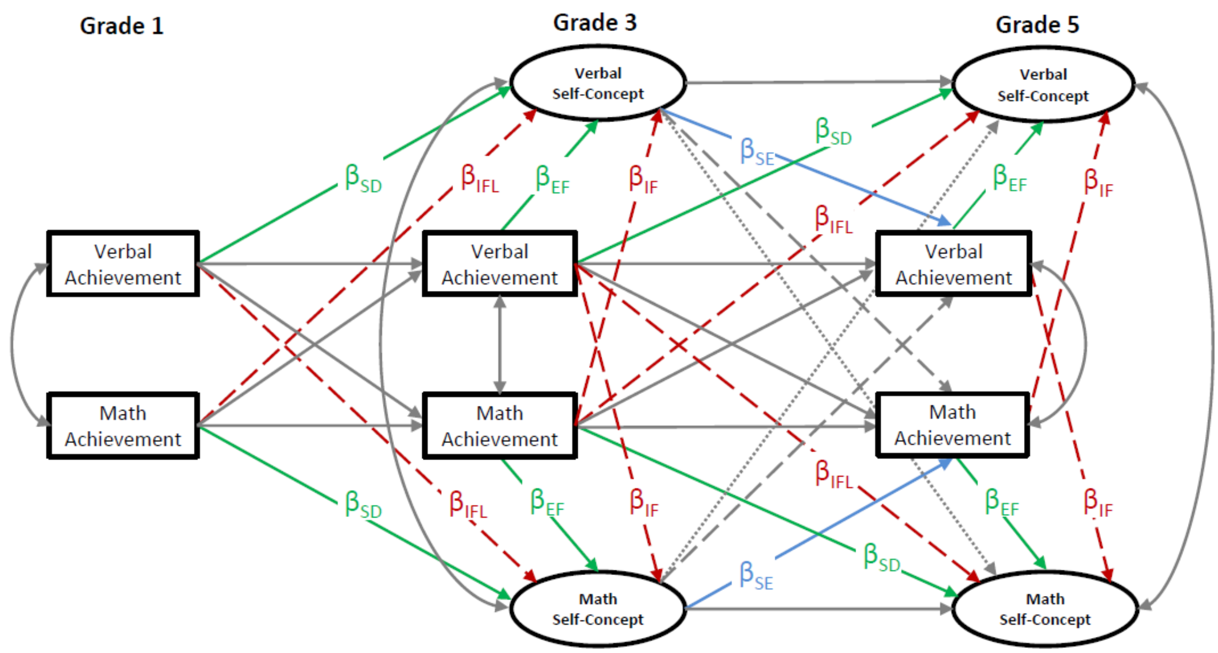


Figure 1. Graphical representation of the reciprocal internal/external frame of reference (RI/EM). *Note*: β_{SD} = skill development effects, longitudinal (green); β_{SE} = self-enhancement effects, longitudinal (blue); β_{EF} = external frame of reference effects, cross-sectional (green); β_{IF} = internal frame of reference effects, cross-sectional (red); β_{IFL} = internal frame of reference effects, longitudinal (red). Predicted positive effects are depicted with continuous lines, negative effects with dashed lines, and null effects with dotted lines. A color version of this figure can be found in the online version of the text. Please note that we present achievement in squares because the latent variables rely on one WLE-score only (for more information on the modeling, see methods). Within-time indicators and across-time correlated uniquenesses are not depicted for reasons of legibility.

In fact, the RI/EM has largely been tested with secondary school students in European countries. For German students, studies have reported positive reciprocal effects of academic self-concept and achievement (grades) within a domain, and negative effects of achievement on subsequent self-concepts across domains (see Möller et al., 2011; Niepel et al., 2014). Similarly, Möller et al. (2014) found positive longitudinal effects of achievement and self-concept within domains using grades and test scores, and negative effects of achievement on subsequent academic self-concepts across domains. However, when taking prior achievement into account, the effects of academic self-concepts of subsequent achievement across domains were near zero. In a sample of Taiwanese students, Chen et al. (2013) longitudinally studied two cohorts of secondary school students and found reciprocal relations between math and Chinese. However, no negative cross-domain effects from prior achievement on subsequent academic self-concepts were present.

At least one recent longitudinal study replicated the typical pattern of I/EM results for primary school students (grades 4 and 5) in Germany (Wolff et al., 2018): strong positive effects from achievement (grades) to the matching self-concepts (social comparison process) and moderate negative effects to the non-matching self-concepts (dimensional comparison process). Moreover, in the longitudinal run, the authors found small positive effects from achievement on matching self-concepts (showing temporal evaluation processes in the subjects), and the effects on nonmatching self-concept proved to be non-significant (signifying that temporal comparison processes in one domain do not affect self-concept formation in other domains). However, their study was confined to grades as an achievement measure, and the generalization of the RI/EM to test scores as an achievement indicator in a primary school setting is sparse. In a very recent study, Weidinger and her colleagues (2019) found no dimensional effects on children's ability self-concept with latent cross-lagged models using a longitudinal data set of German primary school students from grade 2 to 4.

Taken together, the question arises to what extent the RI/EM is falsifiable. We argue that, in cases where the regression weights of the individual effects are close to zero, the assumed effects are not present. Given the size and the representativeness of the sample, such a finding would be strong evidence that certain effects are not meaningful. Furthermore, it might be necessary to introduce additional components into the model to present a holistic view of the processes underlying self-concept formation. For example, Marsh et al. (2017) recently proposed the *Integrated Academic Self-Concept Model*, which also juxtaposes the Big-Fish-Little-Pond-Effect (BFLPE, Marsh & Parker, 1984) with the REM and the I/EM. In brief, the BFLPE focuses on the effect of the average class achievement on students' academic self-concept. However, there is empirical evidence that the BFLPE is strongest for high school students, weaker for students in middle school and college, and weakest for primary school students (Fang et al., 2018). This developmental pattern is reasonable since the effects of transition effects including between class ability tracking usually begin in secondary education. Thereby, the need to investigate the BFLPE with the RI/EM in primary school age sample seems less important and practical in the given setting.

Table 1. Summary of Models on Academic Self-Concept Formation

Acronym	**Model Name**	**Explanation**
REM	Reciprocal Effects Model (Marsh, 1990)	A synthesis of the skill development model (i.e., positive effects of achievement on subsequent academic self-concept within a given domain) and the self-enhancement model (i.e., positive effects of academic self-concept on subsequent achievement within a single domain, see Calsyn and Kenny (1977) for more detail), thus, explaining the mutual dependence between self-

		concept and achievement across time.
I/EM	Internal/External frame of reference model (Marsh, 1986)	The internal frame of reference describes intra-individual or dimensional comparisons of achievement across domains resulting in contrast effects for the self-concepts in the verbal and math domain. The external frame of reference, however, reflects inter-individual or social comparisons within a given domain at a given time point.
RI/EM	Reciprocal Internal/External frame of reference model (Marsh & Köller, 2004)	A combination of REM and the I/EM. That is, the REM (i.e., positive effects of prior academic achievement on subsequent academic self-concept within a domain, and vice versa, Marsh, 1990) is extended across domains. In addition, the I/EM (i.e., negative contrast effects of academic achievement on self-concept across domains, and positive effects of academic self-concept on achievement within one domain) is extended to the longitudinal view.

1.4. The Present Study

Based on the assumptions of the RI/EM, we form the following hypotheses:

1. The REM part of the model will predict (a) positive effects from earlier achievement (verbal and math) on later self-concept within a given domain (*skill development effects*) and (b) positive effects from earlier academic self-concept on later achievement in matching domains (*self-enhancement effects*). In our reading of the literature, the skill development effects seem to be much stronger than the self-enhancement effects. That is, students' current academic self-concept is strongly predicted by their prior achievement within a particular subject, whereas the effect of the self-concept on achievement is less pronounced (see Möller, Pohlmann, Köller, & Marsh, 2009; Valentine, DuBois, & Cooper, 2004). Due to the fact that the effect of academic competence on self-concept is more pronounced and robust than vice versa, we model a directed path instead of a correlation between achievement and self-concept across domains. Please note that the way of modeling affects the interpretation of the coefficients.
2. The I/EM part of the model will predict (a) positive within-domain effects between verbal/math achievement and the corresponding self-concept (*external frame of reference effects*), and (b) negative cross-domain effects between verbal/math achievement on contrasting subsequent academic self-concepts (*internal frame of reference effects*). The I/EM was originally formulated for the cross-sectional case, but in the RI/EM, the predictions are extended to the longitudinal case.

2. Method

2.1. Data Collection and Sample

The data used were obtained from the "Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999" (ECLS-K), a project aimed at assessing educational, physical, and social development among children in the United States over time (Tourangeau et al., 2009).

The ECLS-K represents the efforts of the *National Center for Education Statistics* (NCES) and several education, health, and human services agencies. The ECLS-K data consist of a nationally-representative sample of 21,396 children from diverse ethnic and socioeconomic backgrounds (http://nces.ed.gov/ecls/dataproducts.asp). To establish the sample, a multistage probability design was implemented: the primary sampling units were geographical areas (states or groups of states), the second stage used schools, and the third stage used students. In the present investigation, the total sample size was n = 16,216.

Approximately half of the sample was female (51.2%; n = 10,950). According to self-reports, about 55% of the children were non-Hispanic whites, 15% non-Hispanic blacks, 18% Hispanics, 6% Asians, and 3% reported belonging to other ethnic groups. The children were assessed in seven waves: fall-kindergarten (1998), spring-kindergarten (1999), fall-first grade (1999), spring-first grade (2000), spring-third grade (2002), spring-fifth grade (2004), and spring-eighth grade (2007). The present study used data from the primary schooling period of grades 1, 3, and 5. Due to attrition in the longitudinal study, the amount of missing values varies across waves. For the achievement data, these values represent approximately 13% in grade 1, 12% in grade 3, and 18% in grade 5. For the academic self-concept data, the missing values represent approximately 11% in grade 3, and 19% in grade 5 (for more information on the missing percentage across waves, please see Table 2). The average age at the first measurement occasion was 5.70 years (SD = 0.36). The complete data set, technical manuals, and codebooks are available online from the NCES (Tourangeau et al., 2009).

2.2. Measures

2.2.1. Academic achievement. Students' academic achievement (verbal and math) was assessed with standardized tests. The verbal test measured basic language skills (e.g., letter recognition, sound recognition, decoding multisyllabic words), vocabulary (vocabulary knowledge, receptive vocabulary), and passage comprehension (text interpretation using prior knowledge). In the course of education, the content of the tests changed from basic language skills to more complex comprehension skills in third and fifth grades. The math tests measured students' conceptual knowledge, procedural knowledge, and problem-solving skills. The mathematical content included number sense, properties, and operations, while measurement, geometry, algebra, and basic functions were added later on (for more information, see Tourangeau et al., 2009). The NCES (2005) used a multistage panel review process to develop ECLS-K's verbal and math tests, which were based on the specifications of the NAEP (*National Assessment of Educational Progress*). All test items were exhaustively pilot-tested, and their construct validity was examined with the *Woodcock-McGrew-Werder Mini-Battery of Achievement* (see Pollack, Najarian, Rock, & Atkins-Burnett, 2005; Tourangeau et al., 2009).

The NCES used methods of item-response theory (IRT) to generate adaptive tests that were administered one-on-one to each child in an untimed format. After a brief routing test, each student was administered a test that matched their performance (categorized as low, middle, or high performance). Such a tailored testing procedure minimized the risk of floor and ceiling effects and kept students motivated during the assessment. Technically, the NCES conducted one-parameter IRT (Rasch) analyses for scholastic competence (i.e., verbal and math). To compare students' competence across grades on a common scale, grade-specific test forms were equated using linking items (for details see Tourangeau et al., 2009). For the main analyses, we chose the latest revised version (public 2010/052) of ECLS-K IRT theta scores (verbal and math) as achievement measures, which are criterion-referenced. The IRT reliability coefficients for verbal and math achievement were consistently high for the first, third, and fifth grades (verbal: α = .96/.94/.93; math: .94/.95/.95).

2.2.2. Academic self-concept. Students worked on the *Self-Descriptive Questionnaire II* (Marsh, 1993). In the third and fifth grades, students rated their agreement on a scale from 1 (*not at all true*) to 4 (*very true*) for three items about their verbal and math self-concepts: (a) *I get good grades in reading/math*, (b) *Work in reading/math is easy for me*, and (c) *I am good at reading/math*. Across grades, the self-concept scale showed satisfactory reliability coefficients in terms of Cronbach's alpha (verbal: α = .74/.81; math: α = .81/.86).

2.3. Statistical Analysis

Following data preparation, all analyses in the framework of structural equation modeling (SEM) were conducted using the R package *lavaan* 0.5–23 (Rosseel, 2012). To examine the main research question, we modeled the cross-lagged and stability effects between and for academic self-concept and academic achievement in a longitudinal SEM from grades 1 to 5 (see Figures 1 and 2). Identically-worded items were used as measures of self-concept: as ignoring this correlated uniqueness typically results in positively-biased stability coefficients and a biased model fit (Marsh & Hau, 1996), we, therefore, allowed identically-worded items to correlate (see Marsh et al., 2005).

The maximum likelihood estimator with a robust estimation of standard errors (MLR) was used since the self-concept scales had four response categories and showed no severely skewed distributions (Beauducel & Herzberg, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Academic self-concept was modeled on an item level; the academic achievement scores derived from IRT modeling were included as single-indicator latent variables. All models were estimated using the *Full Information Maximum Likelihood* (FIML) procedure (Schafer &

Graham, 2002), which results in unbiased estimates under the assumption of missingness at random and while retaining statistical power because no observations were deleted (Enders, 2010). To double-check if the way of handling missing data affects the results, we also used *Multiple Imputation by Chained Equations* (MICE, Buuren & Groothuis-Oudshoorn, 2011). Overall, the results of the multiple imputations replicate the original output, both in terms of model fit statistics and parameter estimates. We included the syntax and the results for the additional MI analyses in the online supplement (see Table OS1 and Figure OS1 in the online supplement).

For longitudinal studies, measurement invariance (MI) testing across time points is pivotal for making valid statements about the developmental aspect. Put differently, it is necessary to show that the measurement itself does not change over the course of study (Geiser, 2013). Therefore, we applied longitudinal MI testing (Geiser, 2013; Little, 2013) across time points (grades 1, 3, and 5) to examine the comparability of measurement instruments for academic self-concept and achievement. Subsequently, the model with the most restrictive constraints that was supported by the data was taken as the baseline model to test the RI/EM. We applied a two-index presentation strategy (Hu & Bentler,1999) that combines an absolute fit index such as the *Root Mean Square Error of Approximation* (RMSEA) and an incremental fit index such as the *Comparative Fit Index* (CFI) to evaluate model fit. To make the present analyses transparent and reproducible (Nosek et al., 2015), we have provided all syntax online at the OSF (Center for Open Science, 2018, https://osf.io/tdnfa/).

3. Results

3.1. Descriptive Statistics

Table 1 presents the means, standard deviations, and reliabilities of the verbal/math self-concepts, as well as the verbal/math competence tests from grade 1 to 5. Moreover, the table provides the correlations between manifest scores (lower off-diagonal) and latent variables (upper off-diagonal). All scales showed satisfactory internal consistencies (with Cronbach's α ranging between .74 and 95). There was a steady increase in school performance, which was stronger at the beginning of school education and slightly more pronounced for math. In line with previous work (Marsh, 1989), students on average evaluated their self-concept positively, with a slight decrease during primary school. Verbal and math competencies were stable from grade 1 to 5 (range from r = .69 to .88); the self-concepts were moderately stable from grade 3 and 5 (r = .51/.46). Within a given domain, the link between self-concept and achievement became stronger during primary school (math: from r = .19 to .35, verbal: from r = .25 to .37). Across domains, the correlations between self-concept and achievement ranged from r = -.01 to .20.

Table 2. Means, standard deviations, missingness, reliabilities, and correlations between self-concepts and competence scales.

Variables	*M*	*SD*	Miss[%]	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Verbal self-concept 3	3.22	0.72	11.8	(.74)	.40	.19	.09	.22	.25	.23	.10	.11	.10
2. Verbal self-concept 5	3.02	0.75	19.0	.51	(.81)	.06	.19	.29	.36	.37	.18	.20	.20
3. Math self-concept 3	3.19	0.79	13.5	.26	.10	(.81)	.40	.01	-.01	.01	.14	.19	.19
4. Math self-concept 5	3.02	0.80	17.6	.12	.24	.46	(.86)	.14	.13	.14	.26	.33	.35
5. Verbal competence 1	0.12	0.45	13.7	.25	.32	.01	.15	(.96)	.75	.71	.72	.66	.61
6. Verbal competence 3	0.79	0.31	12.4	.30	.42	-.05	.17	.77	(.94)	.85	.67	.73	.69
7. Verbal competence 5	1.05	0.29	18.1	.29	.44	-.04	.18	.69	.86	(.93)	.63	.72	.73
8. Math competence 1	0.06	0.42	13.8	.12	.23	.15	.27	.73	.69	.65	(.94)	.80	.76
9. Math competence 3	0.73	0.38	13.6	.14	.24	.23	.36	.66	.75	.73	.80	(.95)	.85
10. Math competence 5	1.13	0.40	18.2	.14	.24	.22	.39	.61	.71	.74	.72	.88	(.95)

Note: Miss [%]: percentage of missing data. Reliability estimates (Cronbach's α) are presented in the diagonal in parentheses. The upper off-diagonal gives the manifest correlations, the lower off-diagonal the correlations between latent constructs. All correlations were significant ($p < .01$), except those underlined.

3.2. The Reciprocal Internal/External Model

In a first analytical step, we tested for measurement invariance (MI) across time points (grade 1, 3, and 5), which is a necessary prerequisite for making valid statements in the longitudinal run. The complete results of the MI testing procedure are presented in Appendix A. In summary, even the most restrictive form of longitudinal measurement invariance, strict MI, was supported by the data (n = 16, 216, χ^2 = 2,875.67, df = 110, p <.001, CFI = .975, RMSEA = .039). Accordingly, all subsequent models were specified with invariant factor loadings, intercepts, and residual variances across time points. Moreover, item residuals of identically worded items were allowed to correlate across time points. Below, we first present the reciprocal within-domain effects (i.e., the REM part of the RI/EM), and second the effects of achievement on academic self-concepts across domains (i.e., the I/E frame of reference part of the RI/EM).

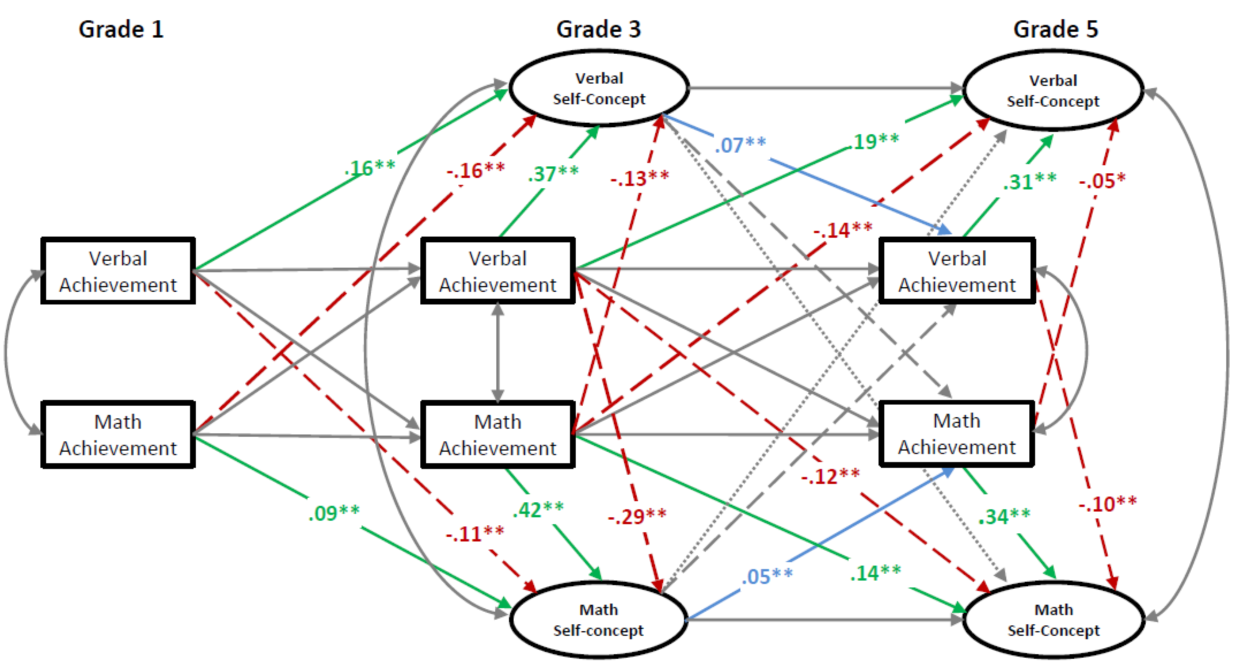


Figure 2. n = 16,216. Results of the reciprocal internal/external frame of reference model in verbal and math domains. *Note*: All coefficients were standardized and significantly different

from zero ($p < .05$). The fit indices of the main models (n = 16, 216, χ^2 = 2,875.67, df = 110, $p < .001$, CFI = .975, RMSEA = .039). Please note that we present achievement in squares because the latent variables rely on one WLE-score only (for more information on the modeling, see methods). Within-time indicators and across-time correlated uniquenesses are not depicted for reasons of legibility.

The REM part of the RI/EM: For the verbal and the math domain, reciprocal relations were evidenced between third and fifth grade (see Figure 2). In more detail, prior academic achievement positively and significantly predicted the individual differences in subsequent academic self-concept beyond the effect of prior self-concept (skill-development effects) for both time periods, for language, β(Vcom1, Vsc3) = .16 and β(Vcom3, Vsc5) = .19; for math, β(Mcom1, Msc3) = .09, β(Mcom3, Msc5) = .14). Also, early academic self-concept positively and significantly predicted the individual differences in later academic achievement over and above the effect of prior academic achievement (self-enhancement effects): for language, β(Vsc3, Vcom5) = .07, and for math, β(Msc3, Mcom5) = .05. To compare the self-enhancement with the skill-development effects, we ran two nested models: the main model with freely estimated regressions vs. a model in which the self-enhancement and the skill development effects are constrained to be equal. The χ^2-difference test revealed a significant deterioration in model fit. That means the effect of prior academic achievement on subsequent academic self-concepts was significantly stronger than in the opposite direction in both domains ($\Delta\chi^2$ (3, N =16,216) = 238.37, p <.001). In terms of underlying concepts (Calsyn & Kenny, 1977), the skill-development paths were more pronounced than the self-enhancement paths.

The I/E part of the RI/EM: The results provided strong support for the internal/external frame of reference effects at both time periods (see Figure 2). More specifically, the internal frame of reference effects were observed cross-sectionally: for language on math, β(Vcom3, Msc3) = -.29 and β(Vcom5, Msc5) = -.10, and for math on language, β(Mcom3, Vsc3) = -.13 and β(Mcom5, Vsc5) = -.05. Similarly, a typical external frame of reference effect was observed at a given time point: for language β(Vcom3, Vsc3) = .37 and β(Vcom5, Vsc 5) = .31, and for math β(Mcom3, Msc3) = .42 and β(Mcom5, Msc5) = .34. In line with the cross-sectional results, we also found longitudinal RI/E effects, that is, negative cross-domain effects of prior achievement on subsequent self-concepts: for language on math, β(Vcom1, Msc3) = -.11, β(Vcom3, Msc5) = -.12, and for math on language, β(Mcom1, Vsc3) = -.16, β(Mcom3, Vsc5) = -.14). Overall, in terms of the underlying concepts (Möller et al., 2009, 2011), the assumptions of the internal/external frame of reference—dimensional and temporal comparisons—were replicated longitudinally among primary school students. Due to the missing of baseline measure of academic self-concept in grade 1, we only reported the skill development effects from academic achievement in grade 1 to self-concept in grade 3. However, to check the effects of prior achievement on subsequent self-concept in the main model, we ran another RI/EM for grades 3 and 5 only (please see Figure OS2 in the online supplement). Overall, the results mirror the ones of the main model reported here.

As a side note, the analyses further revealed close-to-zero correlations between the prior verbal self-concept and subsequent math competence between grades 3 and 5. That is, for language self-concept on math achievement, β(Vsc3, Mcom5) = -.03, and for the prior math self-concept and subsequent verbal competence β(Msc3, Vcom5) = .07. Overall, these coefficients, which can be understood as longitudinal self-enhancement effects, imply that the self-concept does not affect subsequent achievement in an opposite domain. Moreover, the RI/EM showed no effect from prior verbal self-concepts on subsequent math self-concepts between grades 3 and 5 (β = .01), nor between the math self-concept and subsequent verbal self-concepts (β = -.01). The coefficients demonstrate that the self-concepts across domain and time are not related, indicating that temporal comparison processes within one domain do not affect self-concept formation in another domain. Similarly, the cross-sectional analyses also evidenced small correlational coefficients between verbal and math self-concepts in grades 3 (r = .26) and 5 (r = .24).

4. Discussion

In the present study, we examined the integration of the REM and the I/EM into the RI/EM in a US-representative sample of primary school students whom participated in the Early Childhood Longitudinal Study-Kindergarten (ECLS-K, Cohort 1998-1999), which launched a longitudinal investigation into self-concept and scholastic achievement from grade 1 through to 5. More specifically, we investigated the developmental interplay between self-concept and a standardized achievement measure within and across two domains (verbal and math). In terms of underlying theoretical frameworks, we examined to what extent the REM generalizes across the math and verbal domains, and the I/EM generalizes over time.

In line with the assumptions of the REM, we found reciprocal relations in both domains during primary school education: competence and the accompanying self-concept depend on one another over time (see the solid green lines in Figure 1). The mutual dependency does not imply that the effects are equal in size—the effects of academic competence on self-concept (i.e., skill development effects) were significantly stronger than the effects of academic self-concept on competence (i.e., self-enhancement effects). This finding corroborates Wolff et al.'s finding of which positive paths from the change in math/language achievement to the corresponding self-concepts in primary school (2018). The results also add to the existing evidence on the reciprocal relations between multiple academic self-concepts and achievement over time, combined with the predominant effects of skill development effects in primary school-age students (Chapman & Tunmer, 1997; Guay et al., 2003; Helmke & Van Aken, 1995; Skaalvik & Valås, 1999).

The effects of the "classical" I/E frame of reference was also observed in the present study, thus backing up previous cross-sectional research on primary school students (Lohbeck & Möller, 2017; Pinxten et al., 2015). Moreover, in line with the RI/EM, achievement predicted subsequent academic self-concepts in contrasting domains (see the red dashed lines in Figure 1), which replicates previous findings on primary school students (Wolff, Helm, et al., 2018). However, the present finding is in contrast to the recent study of Weidinger et al. (2019) who found no indication for dimensional effects on changes in children's ability self-concepts in the longitudinal run. Comparing the magnitude of the cross-sectional internal frame of reference effects with the longitudinal internal frame of reference effects, it seems that the longitudinal effects do not necessarily decline after two years of cross-lagged effect. This stability is important when considering the long-term effects of dimensional comparison. The only other RI/EM study in a primary school (Wolff, Helm, et al., 2018) found negative but non-significant effects leading from achievement levels and changes to non-corresponding self-concepts, which might have been due to sampling size issues. This evidence is important, since most of the literature on the individual I/EM—both cross-sectional and longitudinal—focuses on secondary school students (Chen et al., 2013; Möller et al., 2009; Möller et al., 2011, 2014; Niepel et al., 2014), omitting the important first years of schooling in self-concept formation (Eccles, Wigfield, Harold, & Blumenfeld, 1991; Marsh, 1985, 1989).

In the present investigation of primary school students, we essentially found the same relations between self-concept and achievement, both cross-sectional and longitudinal, which also indicates that the same cognitive mechanisms are in place. Differently put, primary school students in the third grade in this study already seemed to possess a basic understanding of their own abilities, which allowed them to distinguish between academic domains and do not adhere to a monolithic academic self-concept. This evidence is in line with studies that also indicate that already first-grade students have basic metacognitive skills to evaluate their own competencies and to differentiate between their skills in various areas (Marsh, 1989; Wigfield et al., 1997). However, this finding contradicts other studies that point to an unrealistic overestimation when primary school students assess their self-concept (Dweck, 2002) suggesting a later development of the ability to self-evaluate (e.g., see Craven, McInerney, & Marsh, 2000). The results can also be taken as a cross-cultural validation of the RI/EM. Perhaps due to legal restrictions and a different organization of the school systems, most of the longitudinal studies were carried out in Europe or Asia, with the exception of (Ganley & Lubienski, 2016; Guay et al., 2003). With respect to the RI/EM, to our knowledge only five studies (Chen et al., 2013; Möller et al., 2011; Möller et al., 2014; Niepel et al., 2014; Wolff et al., 2018) were also conducted either in Europe (Belgium, Germany) or Asia (Taiwan). We think that replication studies are vital in science in general and in educational psychology in particular, since the generalizability of results across cultures, school systems, etc., does not only have theoretical implications but often also entails practical educational consequences.

4.1. Limitations and Implications

The present study has certain limitations that should be considered. First, in the literature, the term “causality” is used differently to describe the relationship between two constructs; the labels “cause” and “effect” are sometimes attributed to cross-sectional variables based on theoretical considerations. Some researchers argue that temporal priority indicates causality: variables are affected by temporarily preceding variables in a longitudinal design. In a strict interpretation of causality, causality is given when an effect is produced deliberately by changing or modifying a causing variable in a quasi-experimental setting (see also West & Thoemmes, 2010). In our study, we used the term “effect” in its conventional statistical sense: an effect indicates a longitudinal relationship that is not necessarily causal (see also Marsh et al., 2017; Möller et al., 2014; Niepel et al., 2014; Wolff, Nagy, Helm, & Möller, 2018). However, we infer how students make (social, dimensional, and temporal) comparisons with this limitation in mind. Thus, for strong causal assertions, future research (i.e., experimental, as well as multi-cohort longitudinal design) is needed to scrutinize and to disentangle the different points of comparisons on which students rely in their self-concept formation and scholastic development.

Second, we employed standardized test scores as a measure of academic achievement. In the self-concept literature, it is a common finding that the relations are stronger for grades than test scores (Marsh et al., 2005; Möller et al., 2014; Sewasew et al., 2018), which might be attributed to the fact that students usually do not get any external feedback on a standardized test. Even though it is desirable to juxtapose test scores and grades as achievement measures in order to draw a complete picture of their developmental interplay with academic self-concepts, the current results leave us confident that the derived conclusions are quite robust. Unfortunately, in the present study, academic self-concept was not measured in grade 1. Further studies could employ age-appropriate measures of academic self-concept in young children, for example, by using smileys as response options (Davies & Brember, 1994). Moreover, the beginning of the school career is not yet well investigated—neither cross-culturally, nor across domains (Chiu & Klassen, 2010).

Finally, it is important to mention the limitation of using secondary data to answer competing theoretical models. In the ideal hypothetic-deductive model of science, researchers will collect data with measures that derive from theory-based research problems. In this respect, secondary data sources might deliver biased results because the data were generated with different theoretical models in mind. For example, the item "I think I'm good in math" could be either "I think I'm good in math in comparison to my classmates" if the focus is on social comparison processes or "I think I'm good in math compared to my previous achievements" if the focus is on the developmental aspect. We think the item wording can considerably change the results, which is an aspect that has not been studied in-depth. In the presented study, the academic self-concept items are worded very broadly, which might also explain their widespread use in many areas of psychological research, age groups, and cultures (e.g., Marsh, Ellis, Parada, Richards, & Heubeck, 2005).

Despite the aforementioned caveats, our study tested the typical pattern of the I/EM longitudinally and the REM effects across domains, respectively. We provide further support for the RI/EM assumptions and generalize them to primary school-age students. It is important to stress that the main analyses were conducted after ensuring longitudinal measurement invariance over time (see Appendix A). Furthermore, the present results could be taken as a cross-cultural validation on the RI/EM for a representative sample of US students, since the available studies mostly rely on data from a few (mainly European) countries. The findings have interesting theoretical implications and practical consequences both for psychology and educational research, where only a very small percentage of the publications are replications (see Makel, Plucker, & Hegarty, 2012).

We suggest that academic achievement and self-concept should be fostered simultaneously (see also Marsh et al., 2005). We caution an "either-or" approach—either fostering prior achievement, which affects subsequent academic self-concept (skill development) or fostering prior academic self-concept, which affects subsequent achievement (self-enhancement). For instance, if parents or teachers improve students' academic self-concepts without cultivating their achievement, the student improvements in academic self-concept will not be long-lasting. This fact is particularly relevant for primary school children, because boosting academic self-concept and skills is most influential at this developmental stage. After primary school, the link between self-concept and achievement most likely weakens (Retelsdorf, Köller, & Möller, 2014) due to the fact that the academic self-concepts become more stable and less influenced by prior achievement (Chen et al., 2013). As a result, parental support, academic teaching, and educational programs, in general, should nourish and foster the academic self-concepts and academic skills jointly in the early years of primary schooling (Retelsdorf et al., 2014).

Furthermore, the extended RI/EM results show negative and significant effects of prior academic achievements (e.g., verbal) on contrasting subsequent academic self-concepts (e.g., math), confirming the interplay of dimensional and temporal comparisons processes. If teachers and parents are aware of students' internal and temporal comparisons, the negative effects of academic achievement on the non-matching self-concept can perhaps be mitigated. Therefore, it is important to be aware of and reflect on the devaluating potential of good grades in a specific subject on students' skills in a non-matching subject (Chen et al., 2013; Marsh & Köller, 2004).

4.2. Conclusion

The present study provides strong support for the predictions of the RI/E model, that is, the positive effects between achievement and self-concept in a given domain at a given time point (external frame of reference), and negative effects between achievement and self-concept across domains (internal frame of reference). The I/EM relations were also present longitudinally, thus substantiating the effect of achievement on self-concept across time points (equal to the skill-development part of the REM). Overall, the findings supported the RI/EM in a representative sample of US primary school students using standardized test scores.

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45,* 1–67. https//doi.org/10.18637/jss.v045.i03

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement. *Journal of Educational Psychology*, *69*(2), 136–145. https://doi.org/10.1037/0022-0663.69.2.136

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology*, *67*(3), 279–291. https://doi.org/10.1111/j.2044-8279.1997.tb01244.x

Chen, S.K., Yeh, Y.C., Hwang, F.M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. *Learning and Individual Differences*, *23*, 172–178. https://doi.org/10.1016/j.lindif.2012.07.021

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*(1), 2–17. https://doi.org/10.1016/j.learninstruc.2008.11.002

Craven, R. G., McInerney, V., & Marsh, H. W. (2000). The structure and development of young children’s self-concepts and relation to academic achievement (pp. 1‑20). Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. https://eric.ed.gov/?id=ED443522

Davies, J., & Brember, I. (1994). The reliability and validity of the “smiley” scale. *British Educational Research Journal*, *20*(4), 447–454.

Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. S. Eccles (Eds.), *A Vol. in the educational psychology series. Development of achievement motivation* (pp. 57-88). San Diego, CA, US: Academic Press. http://dx.doi.org/10.1016/B978-012750053-9/50005-X

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1991). Age and gender differences in children’s self- and task perceptions during elementary school. *Child Development*, *64* (3), 830 - 847.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

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

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The Big-Fish-Little-Pond Effect on academic self-concept: A meta-analysis. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01569

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, *47*, 182–193. https://doi.org/10.1016/j.lindif.2016.01.002

Geiser, C. (2013). *Data Analysis with Mplus*. New York: Guilford Press.

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity*, *3*(1), 53–68. https://doi.org/10.1080/13576500342000040

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

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

Harter, S. (2012). The construction of the self: A developmental perspective (First edition). New York: The Guilford Press.

Helmke, A., & Van Aken, G, A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, *87*(4), 624–637. https://doi.org/10.1037/0022-0663.87.4.624

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), 1–55. https://doi.org/10.1080/10705519909540118

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. https://doi.org/10.1016/j.cedpsych.2014.11.002

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model.

Journal of Educational Psychology, *107*(4), 1086–1101. https://doi.org/10.1037/edu0000021

Little, T. (2013). *Longitudinal Structural Equation Modeling*. Guilford Press.

Lohbeck, A., & Möller, J. (2017). Social and dimensional comparison effects on math and reading self-concepts of elementary school children. *Learning and Individual Differences*, *54*, 73–81. https://doi.org/10.1016/j.lindif.2017.01.013

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542. https://doi.org/10.1177/1745691612460688

Marsh, H. W. (1985). Age and sex effects in multiple dimensions of preadolescent self-concept: A replication and extension. *Australian Journal of Psychology*, *37*(2), 197–204. https://doi.org/10.1080/00049538508256398

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi.org/10.1037/0022-0663.81.3.417

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*(4), 646–656. https://doi.org/10.1037/0022-0663.82.4.646

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*(4), 841–860. https://doi.org/10.2307/1163206

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., … Abu-Hilal, M. M. (2015). The internal/external frame of reference model of self-concept and achievement relations: Age-cohort and cross-cultural differences. *American Educational Research Journal*, *52*(1), 168–202. https://doi.org/10.3102/0002831214549453

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. https://doi.org/10.1111/j.1745-6916.2006.00010.x

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*(1), 81–102. https://doi.org/10.1037/1040-3590.17.1.81

Marsh, H. W., Gerlach, E., Trautwein, U., Lüdtke, O., & Brettschneider, W.D. (2007). Longitudinal study of preadolescent sports self-concept and performance: Reciprocal effects and causal ordering. *Child Development*, *78*(6), 1640–1656. https://doi.org/10.1111/j.1467-8624.2007.01094.x

Marsh, H. W., & Hau, K.T. (1996). Assessing goodness of fit. *The Journal of Experimental Education*, *64*(4), 364–390. https://doi.org/10.1080/00220973.1996.10806604

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*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

Marsh, H. W., Hau, K.T., & Kong, C.K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal*, *39*(3), 727–763. https://doi.org/10.3102/00028312039003727

Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of East and West German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology*, *29*(3), 264–282. https://doi.org/10.1016/S0361-476X(03)00034-1

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*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213

Marsh, H. W., Pekrun, R., Murayama, K., Arens, K. A., Parker, P. D., Guo, J., & Dicke, T. (2017). An integrated model of academic self-concept development: academic self-concept, grades, test scores, and tracking over six years. *Developmental Psychology*, 54, 263–280. http://dx.doi.org/10.1037/dev0000393.

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*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: relations to academic self-concept and achievement. *American Educational Research Journal*, *34*(4), 691–720. https://doi.org/10.3102/00028312034004691

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142. https://doi.org/10.1037/0012-1649.38.1.115

Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, *93*(4), 826.

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, *79*(3), 1129–1167. https://doi.org/10.3102/0034654309337522

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 0002831211419649. https://doi.org/10.3102/0002831211419649

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education*, *21*(4), 467–487.

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, *84*(4), 591–611. https://doi.org/10.1111/bjep.12047

Müller-Kalthoff, H., Jansen, M., Schiefer, I. M., Helm, F., Nagy, N., & Möller, J. (2017). A double-edged sword? On the benefit, detriment, and net effect of dimensional comparison on self-concept. *Journal of Educational Psychology*, *109*(7), 1029–1047. https://doi.org/10.1037/edu0000171

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*(4), 1170–1191. https://doi.org/10.1037/a0036307

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Pinxten, M., Wouters, S., Preckel, F., Niepel, C., De Fraine, B., & Verschueren, K. (2015). The formation of academic self-concept in elementary education: A unifying model for external and internal comparisons. *Contemporary Educational Psychology*, *41*, 124–132. https://doi.org/10.1016/j.cedpsych.2014.12.003

Pohlmann, B., & Möller, J. (2009). On the benefit of dimensional comparisons. *Journal of Educational Psychology*, *101*(1), 248–258. https://doi.org/10.1037/a0013151

Pollack, J. M., Najarian, M., Rock, D. A., & Atkins-Burnett, S. (2005). *Early childhood longitudinal study, kindergarten class of 1998-99 (ECLS-K), psychometric report for the fifth grade (*NCES 2006-036). Washington, DC: National Center for Education Statistics.

Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept – testing the reciprocal effects model. *Learning and Instruction*, *29*, 21–30. https://doi.org/10.1016/j.learninstruc.2013.07.004

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Rosseel, Y. (2012). *lavaan*: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36. http://www.jstatsoft.org/v48/i02/

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Sewasew, D., Schroeders, U., Schiefer, I. M., Weirich, S., & Artelt, C. (2018). Development of sex differences in math achievement, self-concept, and interest from grade 5 to 7. *Contemporary Educational Psychology*, *54*, 55–65. https://doi.org/10.1016/j.cedpsych.2018.05.003

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, *67*(2), 135–149. https://doi.org/10.1080/00220979909598349

Tourangeau, K., Nord, C., Le, T., Sorongon, A. G., & Najarian, M. (2009). *Early childhood longitudinal study, kindergarten class of 1998-99 (ECLS-K): Combined user's manual* *for the ECLS-K eighth-grade and K-8 full sample data files and electronic codebooks* (NCES 2009-004). Washington, DC: National Center for Education Statistics.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Vandenberg, R. J., & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2019). Ability self-concept formation in elementary school: No dimensional comparison effects across time. *Developmental Psychology.* Advance online publication. http://dx.doi.org/10.1037/dev0000695

Weinert, F. E. (1999). *Concepts of competence*. OECD: Max Planck Institute for psychological research, Munich, Germany.

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*(1), 18–37. https://doi.org/10.1037/a0015917

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*(3), 451–469. https://doi.org/10.1037/0022-0663.89.3.451

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology,110*(7), 1005-1025. https://doi.org/10.1037/edu0000248

Wolff, F., Nagy, N., Helm, F., & Möller, J. (2018). Testing the internal/external frame of reference model of academic achievement and academic self-concept with open self-concept reports. *Learning and Instruction*, *55*, 58–66. https://doi.org/10.1016/j.learninstruc.2017.09.006

Supplemental Material

Appendix A: Measurement Invariance Testing Across Time

In longitudinal studies, measurement invariance (MI) testing across time points (grades 1, 3, and 5) is pivotal to make valid statements about the developmental aspect (see Geiser, 2013; Little, 2013). In the baseline model, all measurement parameters (i.e., factor loadings, intercepts, and residual variances) were freely estimated across measurement occasions (*configural model*); only the residuals of similarly-worded items were allowed to correlate across time points. Factor means were fixed at the first time point and freely estimated at all other time points. For *metric invariance,* the factor loadings were fixed to be equal across time points. In *scalar invariance* testing, additionally, all intercepts were fixed to be equal across time points. From scalar invariance onwards, factor means were freely estimated, with the exception of the first measurement time point. In the most restrictive form (*strict invariance*), all measurement parameters were equal across time points. To assess whether imposing additional constraints on measurement parameters leads to deterioration in model fit, the differences in the CFI between two consecutive models was evaluated. A difference in the model fit of CFI ≥ .01 is usually considered a serious deterioration (Vandenberg & Lance, 2000).

Table A1. The goodness-of-fit indices for longitudinal MI testing

Model	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	2,123.52	98	. 982	–	.972	.036	–
Metric invariance	2,128.29	102	. 982	.000	.973	.035	.001
Strong invariance	2,165.52	104	. 982	.000	.973	.035	.000
Strict invariance	2,875.67	110	. 975	.007	.966	.039	.004

Note: *n* = 16,216; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Online Supplement

Table OS1: Comparison of model fit statistics and parameter estimates using FIML and MI (Multiple Imputation)

	FIML	MI excluding background variables	MI including background variables
Model Fit Statistics			
χ^2	2,876.29	5,171.66	5,153.10
df	110	110	110
CFI	. 975	.975	.975
TLI	.966	.966	.966
RMSEA	.042	.050	.050
Parameter Estimates			
VAch1 → VSC3	.16	.15	.15
MAch1 → VSC3	-.16	-.18	-.16
MAch1 → MSC3	-.09	-08	-08
VAch1 → MSC3	-.11	-10	-10
VAch3 → VSC3	.37	.37	.37
MAch3 → MSC3	.42	.41	.42
VAch3 → MSC3	-.29	-.27	-.28
MAch3 → VSC3	-.13	-.13	-.11
VAch3 → MSC5	-.12	-.12	-.12
MAch3 → VSC5	-.13	-.12	-.12
VAch3 → VSC5	.19	.15	.15
VSC3 → VAch5	.07	.07	.08
MAch3 → MSC5	.14	.14	.14
MSC3 → MAch5	.05	.05	.05
VAch5 → VSC5	.31	.32	.32
MAch5 → MSC5	.34	.34	.33
VAch5 → MSC5	-.10	-.10	-.10
MAch5 → VSC5	-.05	-.07	-.08

Note: VSC = verbal self-concept, MSC = math self-concept, VAch = verbal achievement, MAch= math achievement. Multiple Imputations are based on 15 data sets.

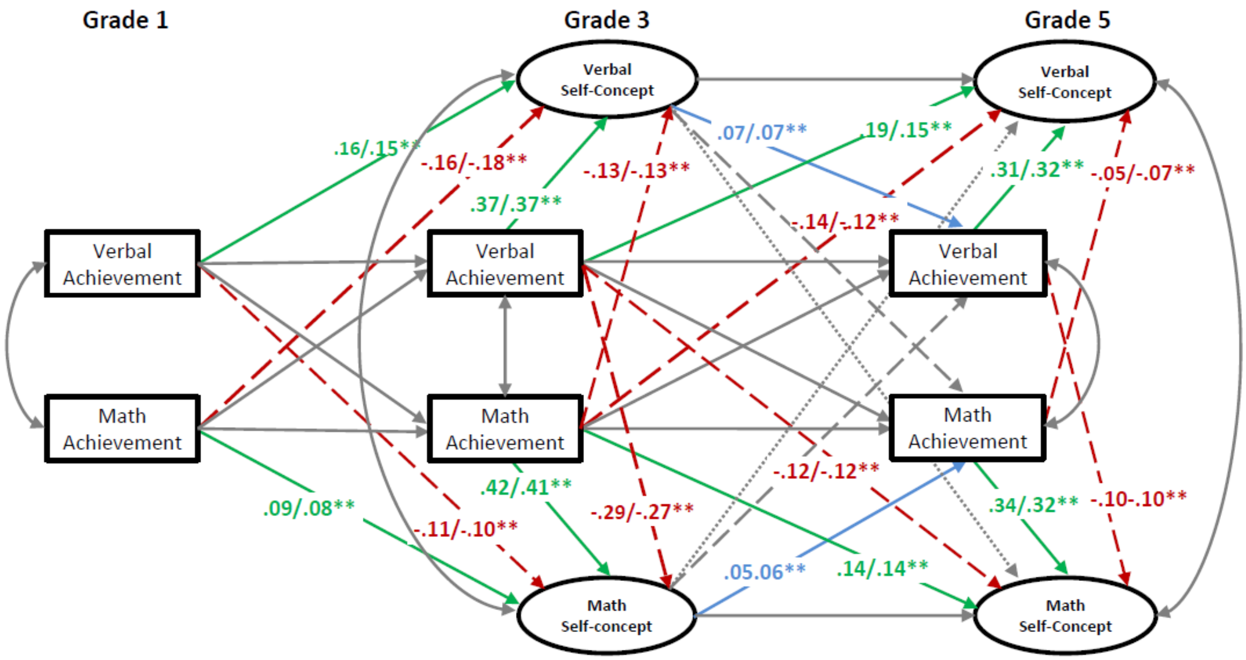


Figure OS1. *n* = 16,216. Results of the reciprocal internal/external frame of reference model in verbal and math domains. *Note*: All coefficients were standardized and significantly different from zero ($p < .05$). The fit indices of the main models based on FIML (n = 16, 216, χ^2 = 2,875.67, df = 110, $p < .001$, CFI = .975, RMSEA = .039), and MI (n = 16, 216, χ^2 = 5,171.66, df = 110, $p < .001$, CFI = .975, RMSEA = .050). Coefficients before slashes are for the main model based on FIML and coefficients after slashes are based on MI. Please note that we present achievement in squares because the latent variables rely on one WLE-score only (for more information on the modeling, see methods). Within-time indicators and across-time correlated uniquenesses are not depicted for reasons of legibility.

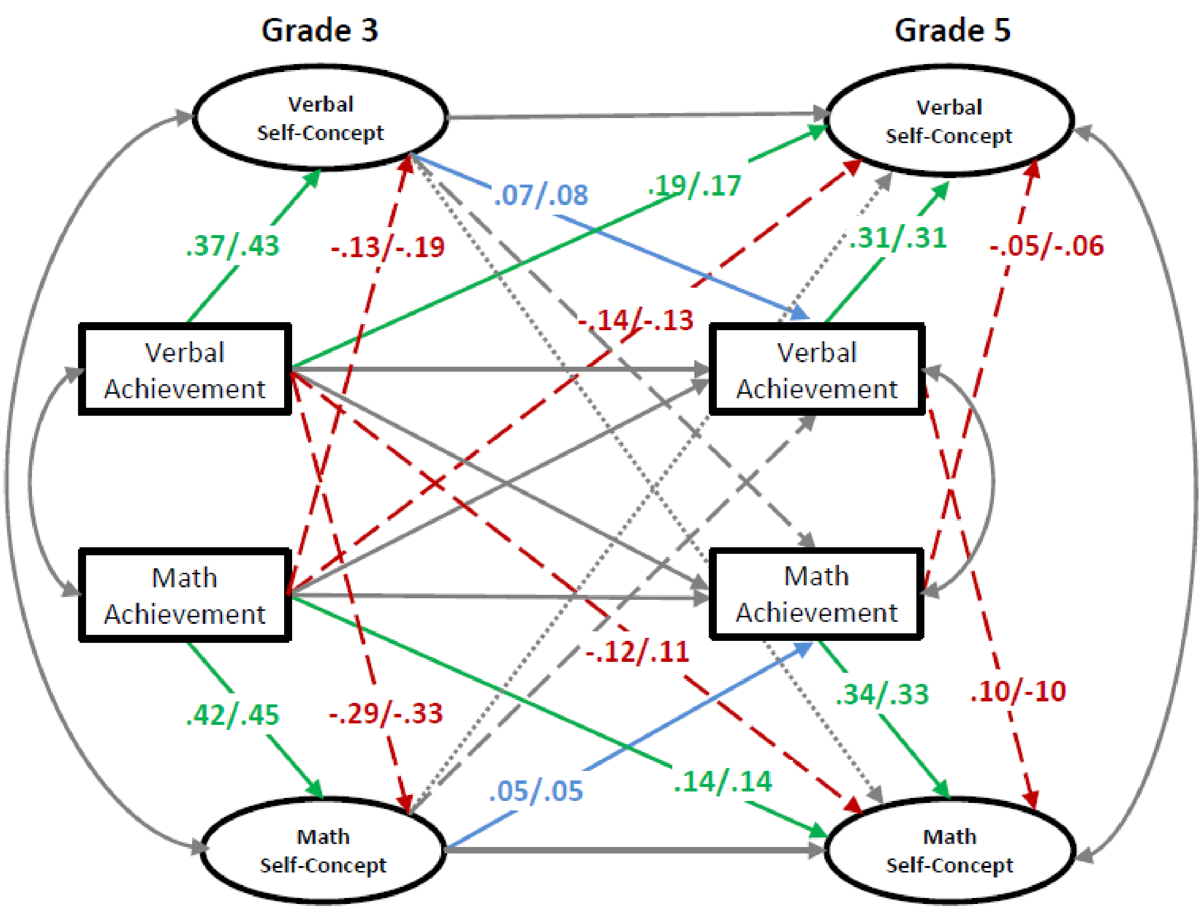


Figure OS2. Results of the reciprocal internal/external frame of reference model in verbal and math domains. *Note*: n = 16,216. Three time points: χ^2 = 2,875.67, df = 110, p < .001, CFI = .975, RMSEA = .039; Two time points: χ^2 = 2,331.91, df = 86, p <.001, CFI = .973. All coefficients were standardized and significantly different from zero (p < .05). Coefficients before slashes are for the main model including all variables together and coefficients after slashes are the second model after excluding Grade 1 achievements. Please note that we present achievement in squares because the latent variables rely on one WLE-score only (for more information on the modeling, see methods). Within-time indicators and across-time correlated uniquenesses are not depicted for reasons of legibility.

The Developmental Interplay of Students' Reading Self-Concept and Reading Competence: Examining Reciprocal Relations and Ethnic-Background Patterns

Daniel Sewasew[a]

Lynne Sanford Koester[b]

[a] *Bamberg Graduate School of Social Sciences (BAGSS), University of Bamberg, Germany*

[b] *Department of Psychology, University of Montana-Missoula, USA*

The data utilized for this study are obtained from the "Early Childhood Longitudinal Study, Kindergarten (KCLS-K)", a project aimed at assessing educational, physical, and social development among children across over time (Tourangeau, Nord, Le, Sorongon, & Najarian, 2009). All ECLS-K, Kindergarten-Eighth Grade Public-use File are available from the National Center for Educational Statistics website (http://nces.ed.gov/ecls/dataproducts.asp).

Correspondence concerning this article should be addressed to Daniel Sewasew, Bamberg Graduate School of Social Sciences, University of Bamberg, Feldkirchenstraße 21, 96045 Bamberg, Germany, E-mail: sewasewdaniel@gmail.com, Phone: +49 (0)951/863-3617

Abstract

Despite the fact that reading self-concept and reading competence are considerably correlated, the reciprocal relation of these variables and the associated ethnic-background patterns are still a fundamental question with significant theoretical implications and practical consequences. Using a Reciprocal Effects Model (REM), we analyzed primary school-aged children in the United States who took part in a representative longitudinal data collection: the Early Childhood Longitudinal Study-Kindergarten (ECLS-K). Reading self-concept and reading competence were found to be reciprocally related for primary school-aged children; however, the influence of prior reading competence on subsequent reading self-concept (the skill development part of REM) was significantly stronger than *vice versa*. The REM of reading competence and reading self-concept was moderated by children's ethnic-background (evidenced only for White-majority students). While longitudinally the descriptive results revealed comparable reading self-concept across the ethnic-backgrounds, there was a larger ethnic disparity for White and Asian over Black and Hispanic family students in reading competence. The study stresses the importance of implementing intervention-based schemes (e.g., learning to read rather than reading to learn) for reading deficits at the early stages of schooling.

Keywords: Reading self-concept, Reading competence, Ethnic-background, Reciprocal effects model, Primary school

Footnote (introduction section)

[1] Ethnic and/or migration background/origin are the two most often overlapping and slippery terms in the social and educational research literature. Studies in Europe have consistently used the term migrant to denote students' parental origin as a *native* (White majorities), or an *immigrant* (ethnic-minorities) student (i.e., parental roots other than host country). In North America, particularly in the U.S., the ethnic/racial origin is most often used in the literature to denote parental lineage of White Europeans (*majorities*) and Black Americans/Hispanic Latinos/Asian (*minorities*), as historically the majority of the population has an immigrant background. However, in the present study of the current school populations, we used these terms interchangeably depending on the site of the study to refer to *ethnic-majority* students (White/native) vs. *ethnic-minority* students (Black/Hispanic/Asian/immigrant) based on parental and grandparental origin.

The Developmental Interplay of Students' Reading Self-Concept and Reading Competence: Examining Reciprocal Relations and Ethnic-Background Patterns

1. Introduction

Reading ability plays a key role in students' academic success, as academic competence is highly associated with students' level of reading proficiency. Wide differences are noticed in reading competence among school-aged children, in part attributed to the level of cognitive competences (e.g., IQ and phoneme awareness; Warmington & Hulme, 2012). In addition, studies have claimed that reading self-concept is linked with the progress of reading ability, apart from intellectual abilities (Morgan & Fuchs, 2007). Yet theories are diverse regarding the developmental nature of reading competence and self-concept relationships. For example, prior reading self-concept may be a primary determinant/predicator of current reading competence but not vice versa; prior reading competence may be a major determinant/predicator of current reading self-concept but not vice versa; and reading competence and reading self-concept may be reciprocally related across time.

Literature has recently emerged on the reciprocal relationship of reading competence and self-concept (for an overview, see Möller, Retelsdorf, Köller, & Marsh, 2011; Möller, Zimmermann, & Köller, 2014; Retelsdorf, Köller, & Möller, 2014). However, despite studies increasingly endorsing this relation, the results concerning the direction, strength, and significance have been mixed. Therefore, the important question in the long-standing debate between determinants of reading competence vs. reading self-concept is which pathways are more predominant, and this has both theoretical implications and practical consequences in interventions.

The generalizability of reciprocal relations in reading competence and self-concept among primary school sample grades is still open to question. This might be because the academic competence and self-concept relations are domain-specific (Marsh & Craven, 2006); they are often overshadowed by the intense research focus on a certain subject (Math) or specific school sample (secondary school-aged students); research designs have been primarily cross-sectional (Chapman, Tunmer, & Prochnow, 2000; Morgan & Fuchs, 2007); and inappropriate measures of reading self-concept have been employed (Chapman & Tunmer, 1995). In this context, the limited existing studies have not taken the latest methodological advances (e.g., ensuring longitudinal and multi-group measurement invariance testing, specifying competence measures at the latent level using item response theory, IRT) into consideration. It could be due to this weakness that there is not yet a clear depiction of the reciprocal relationship of the competence self-concept in the domain of reading (Retelsdorf et al., 2014), particularly in primary school-age children.

The relationship between reading competence disparity and ethnicity has gained research priority as the ethnic-background of students' family impacts reading competence. For example, large international surveys (including the USA), as well as country-specific studies, reported that children from ethnic-majority families (Whites) consistently exhibited stronger reading skills than ethnic-minority students (i.e., Blacks and Hispanics; OECD, 2010; Powell, Son, File, & Froiland, 2012). However, some ethnic-minority groups (i.e., Asian students) often scored stronger in reading competence than other minority students (Black and Hispanic; Peng & Wright, 1994). Diverse plausible reasons are forwarded for the reading competence disparities including family socio-economic status (SES), home and school situations, and individual characteristics (e.g., sex and ethnicity; see Duncan & Magnuson, 2005; Singh, Chang, & Dika, 2010). Still, some have suggested the perceived stereotype threat—ethnic-minority students might have experienced negative reading competence results (e.g., lesser marks and less engagement in reading) because they are loaded by the understanding of endorsing categorizes questioning their academic competence (Steele, Spencer, & Aronson, 2002). Others have suggested the perceived competence (i.e., reading self-concept) as one explanatory reason for ethnic disparities in reading competence; however, differences in reading self-concept across groups do not entirely reveal inequalities in reading competence (Ogbu, 1993; Rodriguez, 2014). Collectively, studies relating reading self-concept to ethnicity and reading competence have not revealed dependable findings and have mostly been inconsistent and unsettled. Thus, it is the goal of the current study to better comprehend the nature of the relation between reading self-concept and reading competence longitudinally, especially in primary school years where studies are scant. Overall, we aimed to test the reciprocal relation between reading self-concept and competence and its potential differentiation regarding ethnic origin from data in three waves of a large, longitudinal study conducted on a representative U.S. primary school sample (i.e., grades 1 to 5).

1.1. Reciprocal Relations Between Reading Self-Concept and Reading Competence

Reading self-concept is defined as students' perceptions of their reading competence and the degree to which the students view reading as a valuable skill (Chapman et al., 2000), whereas reading competence refers to students' level of ability, mastery, or confidence in a specific content area (Schiefele, Schaffner, Moller, & Wigfield, 2012). Success in reading competence is closely tied to reading self-concept, and self-concept formation is largely determined by the opinion of significant others, concrete feedback and causal attribution (Shavelson, Hubner, & Stanton, 1976). However, one unsettled concern is how the reading self-concept and competence relations are temporally ordered (i.e., how prior reading self-concept predicates subsequent reading competence and vice versa). Several theories have been proposed to address this temporal ordering in the development of reading self-concept and competence relationship.

First is the *skill development* hypothesis (Calsyn & Kenny, 1977) in which reading competence is regarded as impacting reading self-concept (e.g., students feel good about themselves because they do well in reading), whereas reading self-concept does not affect reading competence. When longitudinal investigations are considered to check this assumption, i.e., prior reading competence on subsequent reading self-concept, a significant effect is expected; however, in the case of prior reading self-concept and subsequent reading competence, a non-significant effect would be expected (Huang, 2011). The skill development effect occurs as the result of students' judgement of their competence as superior in comparison with their classmates (social comparison; Festinger, 1954), with their prior competence (temporal comparison; Albert, 1977), and positively affects their current reading self-concept in the same subject, and vice versa. A number of studies have reported such support for this model: Skaalvik and Valås (1999) assessed Norwegian primary and middle school students using a two-wave, two-domains (i.e., verbal and math) design to examine this relation in three cohorts. They found reading achievement significantly affected subsequent reading self-concept in two cohorts, whereas prior reading self-concept did not affect subsequent reading achievement. Similarly, Möller et al. (2014) in their longitudinal design and sophisticated analyses model (i.e., the reciprocal internal/external frame of reference model, RI/EM) with a German secondary school sample found a unidirectional effect; reading achievement significantly affected reading self-concept and this result was replicated in two achievement measures (i.e., test scores and grades).

The reading self-concept failed to significantly predict reading achievement, using either grades or test scores.

Second, in the *self-enhancement* hypothesis (Calsyn & Kenny, 1977), reading self-concept is considered to be a major precondition for reading competence (e.g., students do well in reading because they feel good about themselves); conversely, reading competence is not taken as a main factor of reading self-concept. This could occur with more commitment and work in domain-specific activities, which may then result in higher competence (Marsh & Yeung, 1997). Consequently, the optimal effective approach to progress in reading competence is by enriching reading self-concept. The self-enhancement effects come about as the result of students' judgement about their reading self-concept as superior in comparison with their classmates (social comparison; Festinger, 1954), with their prior reading self-concept (temporal comparison; Albert, 1977), and may positively affect their current reading performance in the same subject, and vice versa. The empirical support for the self-enhancement model is not abundant. Early work by (Marsh, 1990) analyzed reading self-concept and achievement based on the data obtained from older children using standardized tests (i.e., IQ, vocabulary and reading comprehension), self-reported Grade Point Average (GPA), and reading self-concept. The author reported that reading self-concept influences subsequent reading achievement, whereas the effect of prior reading achievement on subsequent reading self-concept was not supported. In another study (Valentine, DuBois, & Cooper, 2004), meta-analyses documented a small advantage indicating a favorable influence of positive reading self-concept on reading competence. However, the crucial point here is that the extent of path effects is not that relevant when examining the reciprocal relations between reading self-concept and competence. That is, even though the relationship is best seen as reciprocal, the most pronounced significant path was the effect of academic competence on self-concept, thus the empirically and theoretically important question arises whether academic self-concept also has significant effects on subsequent academic achievement (Marsh & Craven, 2006).

Finally, the *integrated* hypothesis, i.e., Reciprocal Effects Model (REM), indicates that prior reading self-concept affects subsequent reading competence and that prior reading competence impacts subsequent reading self-concept (e.g., students do well in reading as well as feel good about themselves in combination; Marsh & Craven, 2006). The mutual effects occur when students make the social and temporal comparisons reciprocally, i.e., they relate their current performance in a reading domain with their own prior performance in the same domain and evaluate their reading self-concept, and vice versa. Accordingly, effective intervention should simultaneously impact reading competence and reading self-concept (Marsh & Martin, 2011). Only a few studies provide support for this model in a sample of primary school children: Guay, Marsh, and Boivin (2003) analyzed young French children (grades 2 - 4) using a three-wave design for the testing of three cohorts regarding academic grades (i.e., reading, writing, and math) and competence belief (i.e., self-concept). They found support for the reciprocal effects model between reading performance and self-concept, and this result was invariant across the three cohorts. Another recent study using Norwegian first graders at risk of reading difficulties revealed the bidirectional relationships between standardized reading tests and self-concept (Walgermo, Foldnes, Uppstad, & Solheim, 2018). More studies are emerging in secondary school samples (German and Taiwan, e.g., Chen et al., 2013; Möller et al., 2014; Niepel et al., 2014) analyzing the REM of teacher-assigned grades (i.e., German, English, and Chinese subjects) and reading self-concept. These have evidenced reciprocal relations between reading competence and self-concept and vice versa. However, a closer look at all of these studies shows imbalanced reciprocal relations between reading competence and self-concept (i.e., strong skill development effects). An additional study (Retelsdorf et al., 2014) with a German secondary school sample student using standardized reading test scores and reading self-concept with rigorous statistical analyses found total effects of reading competence on self-concept along all paths of analyses; however, reciprocal effects were found only at the beginning of secondary school.

Intuitively, we might conclude that there is little agreement on the temporal ordering of reading self-concept and competence. Some have claimed that development may inhibit one from detecting a bidirectional relationship between achievement and academic self-concept in the younger age children (Chapman & Tunmer, 1997; Skaalvik & Skaalvik, 2004). Meanwhile, self-concept may be reliant on prior achievement, not vice versa (Skaalvik & Valås, 1999). This may signify a skill development effect for younger children, and a reciprocal effect for older age children (Fraine, Damme, & Onghena, 2007). However, Guay et al. (2003) found the reciprocal effects model for primary school pupils and corroborated that as children become older, the rating of academic self-concept becomes more reliable and more stable. Some others claimed that methodological differences might describe the discrepancies originating among earlier studies, such as differences in a) analysis approach, manifest indicator vs. latent variable modeling (see Möller et al., 2014; Retelsdorf et al., 2014); b) order of measured variables, concurrently vs. subsequently (see Chen et al., 2013; Möller et al., 2011, 2014; Niepel et al., 2014); or c) time gap between waves, short vs. large (Chen et al., 2013; Niepel et al., 2014). Still, some others argued that we cannot ignore the impact of school systems on the divergent result of competence and self-concept relationships (D'hondt, Van Praag, Van Houtte, & Stevens, 2016). For example, effects may stem from arrangement of early childhood education and associated experiences, private vs. public school (e.g., learning resource availabilities), school autonomy (e.g., right to change course syllabus), school tracking system (academic vs. non-academic), and other treatments (e.g., class size, teaching methods, teacher performance pay). Overall, a differing relation between reading self-concept and competence for primary school students is confirmed by a wide review of the literature.

1.2. The Role of Ethnic-Background in the Relationships Between Reading Self-Concept and Reading Competence

Academic accomplishment is a significant factor in the adaptation to school and in later life success for children of varied ethnic origins. Ethnicity as an issue in academic self-concept and achievement has been extensively examined but the relationships remain undecided. Studies speculate that lower academic competence among ethnic-minority students is associated with a lower self-assessment in these students (see Singh et al., 2010). As stated by theories of self-evaluation, evidence about competence and comparisons with others is handled into self-judgements. The social comparison theory asserts that social comparisons following positive ability input will prompt a more positive self-concept and the other way around (Festinger, 1954). The process of self-evaluation could combine the feedback from significant others (i.e., symbolic interaction; Mead, 1934).

However, in contrast to the self-evaluation theories (i.e., lower self-concept based on lower competence), ethnic-minority students (e.g., Black, Hispanic and Asian) have largely been found to have equal or higher academic self-concept compared to White majority students (Hall, 2014; Singh et al., 2010). The phenomenon of positive self-evaluation of ethnic-minority students despite underachievement is known as the aspiration-achievement paradox. For example, in reading self-concept, ethnic-minority students (Black and Hispanic) rather astonishingly show more positive reading self-concept than White majority students (Bécares & Priest, 2015). Conversely, in the U.S., in reading and math tests White students score higher on average than all other ethnic groups, predominantly when compared to Black and Hispanic students (Bécares & Priest, 2015). In a German sample, students' ethnic-background appears to play a significant role in predicting their reading competence as ethnic-minority students show significantly lower reading competence than their ethnic-majority peers (see Kigel, McElvany, & Becker, 2015; Miyamoto, Pfost, & Artelt, 2017).

Several plausible explanations have been suggested linked to the absence of a strong association between reading self-concept and competence for ethnic-minority students. According to the perceived discrimination approach, when students are a member of a group that is unfavorably stereotyped in explicit subjects (e.g., reading), those students may identify or reconceive their reading self-concept to limit reading competence from their self-evaluation process (Steele, 1997). Eventually, this perceived stereotype threat results in less engagement on learning tasks and students experiencing test anxiety (Steele, 1997). On the other hand, in the school context, discrimination could stem from teachers' treatment of students, as ethnic-minority groups' teachers are often dissimilar. For example, Hispanic and Black students reported being controlled more punitively by teachers and getting lower grades because of ethnicity (Fisher, Wallace, & Fenton, 2000). The unfair treatment by teachers might lead students to dislike subjects, devalue school, distrust teachers, and perceive teachers as uncaring and useless (Rosenbloom & Way, 2004). An additional possibility includes the home situation, particularly language barriers; for example, speaking a different language in the home has a negative effect on minority students' reading competence and self-concept (August, Shanahan, & Escamilla, 2009; Hall, 2014). Overall, these views are not irreconcilable but can supplement each other in providing a justification of why ethnic-minorities have greater academic aspirations than ethnic-majorities.

With respect to the longitudinal study perspective, the role of reading self-concept in elucidating ethnic gaps between students as opposed to partialling out ethnic-background in reading competence remains undecided. For example, the reciprocal relations between reading competence and self-concept are not restricted to a specific subgroup, that is, whether students have or do not have a migration background (see Möller et al., 2014); however, reciprocal relations are only evidenced in ethnic-majority secondary school students in some studies (see Miyamoto et al., 2017; Schaffner, Philipp, & Schiefele, 2016). Considering the competence gaps highlighted before, the hypothesis of a reciprocal positive relationship between reading competence and self-concept makes a lower reading self-concept in students with ethnic-minority students conceivable. Therefore, it is vital to study the influence of motivational indicators such as reading self-concept on reading competence (and vice versa) for students with different ethnic origins over time.

1.3. The Present Investigation

Several research questions and hypotheses directed this study, as follows:

1. Are there reciprocal relations among U.S. students' reading self-concept and competence during primary school age (grades 1 to 5)?
    a. Hypothesis: Based on past studies, we expect to find reciprocal positive relations between reading self-concept and competence among primary school students (Guay et al., 2003; Walgermo et al., 2018).
2. Do the reciprocal relations differ by ethnic-backgrounds (White/Black/Hispanic/Asian) of students?

   a. We do not expect relations to differ by ethnic-background (Möller et al., 2014).

3. Are there ethnic differences in the U.S. students' reading self-concept and competence during primary school age (grades 1 to 5)? How does the magnitude of such differences compare to one another across time points?
   a. We predict that there will be large, statistically significant ethnic differences in reading self-concept and competence at all grade levels (Bécares & Priest, 2015; Steele, 1997).
   b. The ethnic differences in reading self-concept will be comparable to those in reading competence at each grade (Miyamoto et al., 2017; Salikutluk, 2016; Schaffner et al., 2016). However, we have no hypothesis with respect to the relative size of ethnic gaps in comparison to each other.

2. Method

2.1. Data Collection and Sample

Participants in this study were obtained from the "Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K)", a project aimed at assessing educational, physical, and social development among children in the U.S. over time (Tourangeau, Nord, Le, Sorongon, & Najarian, 2009). The ECLS-K is sponsored by the U.S. Department of Education, NCES (National Center for Education Statistics). The ECLS-K data included a nationally representative sample of over 21,396 children from diverse ethnic and socioeconomic backgrounds. The ECLS-K used multistage probability design; the primary sampling units were geographical areas (states or group of states), the second stage was a school within the primary units, and the third stage was students within the schools.

Of the total students, 51.1% (n = 10, 950) were males and 48.8% (n =10, 446) females; about 55% of children self-reported non-Hispanic White, 15% non-Hispanic Black, 18% Hispanic, 6% Asian, and 3% reported being another ethnicity. The children were assessed in seven waves: fall-kindergarten (1998), spring-kindergarten (1999), fall-first grade (1999), spring-first grade (2000), spring-third grade (2002), spring-fifth grade (2004), and spring-eighth grade (2007). The present study used data from the primary school period, grades 1, 3 and 5. Owing to attrition in the longitudinal study, the amount of missing values varies across waves. For the competence data, these values represent approximately 13% in grade 1, 12% in grade 3, and 18% in grade 5. For the academic self-concept data, the missing values represent approximately 11% in grade 3, and 19% in grade 5 (for more information on the missing percentage across waves, please see the bottom of Table 2). The average child's age during the first grade was 6.53 years (SD = 0.34), whereas at grade 3 and 5, 8.67 (SD = 0.33) and 10.59 (0.29) years respectively. All waves of data are available from the NCES, as well as a technical manual describing the use of the data, sampling plans, and weights (http://nces.ed.gov/ecls/dataproducts.asp).

2.2. Measures

2.2.1. Reading competence. Students' reading skill was assessed using standardized tests. The reading test measured basic skills comprising letter recognition, sound recognition, decoding multisyllabic words, sight vocabulary, vocabulary knowledge, and passage comprehension. The coverage of the tests changed from basic reading skills to comprehension in the third-grade tests and afterwards—*initial understanding*, *developing an interpretation*, *personal reflection and response*, *demonstrating a critical stance* (for more information, see Tourangeau et al., 2009). The ECLS-K reading test items were developed with the assistance of several sources and specialists, including national and state benchmarks (e.g., NAEP, *National*

Assessment of Educational Progress), existing state and commercial assessments, teachers, elementary curriculum and content specialists, and multicultural experts. Test items were exhaustively pilot-tested and their construct validity was initially proven by confirming that student performance on the items correlated with their performance on the well-known Woodcock-McGrew-Werder mini-battery of achievement (see Pollack, Najarian, Rock, & Atkins-Burnett, 2005; Tourangeau et al., 2009).

The NCES applied techniques of IRT to develop adaptive tests that were administered exclusive with each individual child in an untimed arrangement. Reading competence assessments were individually administered with trained examiners who read the questions to the children, but the children read the response options themselves. Before the actual assessment, children were given separate routing assessment forms (twelve to twenty items) to determine the level of their reading ability (Tourangeau et al., 2009). Following the mark of the individual routing procedures, the trained examiner then allocated a low or high another phase-level arrangement of the reading test. This procedure exposed students to tests that were within the limits of and just beyond their optimal challenges range. Such test procedures reduce the chance of floor and ceiling effects through letting all students to have similar possibility to score improvements on the test, irrespective of their preliminary attainment. NCES performed one-parameter IRT (Rasch) analyses for direct cognitive assessments (i.e., reading, math, and science). To compare students' competence across grades on a common scale, grade-specific test forms were equated using linking items (for details, see Tourangeau et al., 2009). NCES regards use of the IRT scores as the most fitting metric for longitudinal modeling, as these scores can be compared across different test form measures and across different grades. The IRT theta scores estimated students' reading ability in a continuous scale across different time points. For the main model analyses (REM), the ECLS-K IRT theta scores, which are criterion-referenced, corrected, the latest revised version (public 2010/052) and provided by the ECLS-K (C4R4RTHT_R, C5R4RTHT_R, and C6R4RTHT_R for 1th, 3rd and 5th grades respectively) were used as the reading competence measure. The IRT reliability figures were consistently high for first, third and fifth grades (α = . 96/. 94/. 93; Tourangeau et al., 2009).

2.2.2. Reading self-concept. Students filled an individually administrated Self-Descriptive Questionnaire (SDQ II; Marsh, 1992). The items were adapted with the permission from the SDQ II: a theoretical and empirical basis for the measurement of multi-dimensions of academic self-concept, and designed for children in late elementary and middle school (Tourangeau et al., 2004). In the third and fifth grades, students responded to three items about their reading self-concept. Students responded on a scale from 1 to 4, with 1 being "*not at all true*", 2 being "*a little bit true*", 3 being "*mostly true*" and 4 being "*very true.*" The items were: *I get good grades in reading, Work in reading is easy for me, I am good at reading*. Trained test administrators read the SDQ questions aloud and provided the student a few seconds to record their answers so as to avoid reading ability influencing the answers. The assessors did not stare at the students' answers as they were recording them, thus lessening the possibility that the students' replies were influenced by an attempt to please the examiner. An individual student completed the SDQ in roughly five minutes. The items are comparable to the items previously used to measure the perceptions of competence in reading (Chapman & Tunmer, 1995), and the cognitive-evaluative aspect of students' reading self-concept (Marsh, 1990; Möller & Pohlmann, 2010). In grade one, reading self-concept items were not administrated, as the wording of the items were not plausible for young children. Across grades, the self-concept scale showed satisfactory reliability coefficients in terms of Cronbach's alpha (α = .74/.81).

2.2.3. Socioeconomic status (SES). A continuous SES measure included in the dataset was used to control for SES. In the ECLS-K dataset, SES was computed at the household level using data for the set of parents who completed the parental interview in the respective waves. The components used to create the SES were mother's and father's education and occupation, and household income. A one-factor SES composite score was used for the REM analyses.

2.3. Statistical Analysis

Following data preparation, all analyses were conducted using the "*lavaan*" R statistical package, version 0.5–23 (Rosseel, 2012). We used three waves and two-year autoregressive cross-lagged panel design Structural Equation Modeling (SEM) for the estimation of the REM. The maximum likelihood estimator with a robust estimation of standard errors (MLR) was used due to the large sample involved, the self-concept scales had four response categories and showed no severely skewed distributions (Beauducel & Herzberg, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Reading self-concept was modeled on an item level. A latent variable longitudinal SEM was analyzed to determine the magnitude of the effects of reading competence on subsequent self-concept and the effects of reading self-concept on subsequent reading competence. A suitable standardized longitudinal weight (i.e., C456CW0) was used, therefore samples are nationally representative of the U.S. (Tourangeau et al., 2009). Also, the inclusion of longitudinal sampling weight in all analyses allowed consideration of the clustering of students within schools. We utilized the MLR estimator instead of the common ML estimation so as to account for the sample weights; the lavaan algorism takes into account the dependency of observations owing to clustering sampling during weighted analysis (Rosseel, 2012). Lavaan handles missing data through the Full Information Maximum Likelihood (FIML) procedure. FIML considers about measurement points or waves for a student are related over time;

therefore, this estimator uses all existing data from the prior to later measurement events to estimate model parameter and residual values. FIML is a strongly recommended procedure for handling missing data (Enders, 2010).

2.3.1. Measurement invariance testing across time and ethnic-background. For longitudinal investigations, measurement invariance (MI) analysis over measurement points/waves is essential to claim valid inferences about age-related changes. It is indispensable to ensure how well the individual items in each measurement or scale hang together overtime (Geiser, 2013). For this reason, we employed longitudinal MI testing (Geiser, 2013; Little, 2013) across time points (grades 1, 3, and 5) to test the comparability of measurement instruments for reading competence, self-concept (see further discussion in supplemental materials appendix A & B, and Tables S1 & S2; Vandenberg & Lance, 2000). For invariance testing across ethnic origin groups, we ran the most restrictive form of longitudinal measurement invariance we could establish — a series of increasingly restrictive models within the framework of multiple-group confirmatory factor analysis (Schroeders & Gnambs, 2018).

2.3.2. Reciprocal effects model. For the first research question, structural equation modeling was used. Overall, we specified two separate series of models for each of the reading standardized test scores and self-concept measures. First, we modeled *cross-lagged* and *stability* effects of individual differences in reading measures and in self-concept from grade 1 to 5 for the overall sample (see Figure 1, Model 1). Second, to understand the ethnic-background groups' effect, we examined a multi-group REM for reading self-concept and competence (see Figure 2, Model 2). For the ethnic origin specific investigation, the critical test is whether the regression coefficients leading from earlier reading self-concept to later reading competence and from earlier reading competence to later reading self-concept across ethnic-background groups are identical. That is, if constraining the cross-lagged paths to equality across ethnic-background groups leads to significant deterioration in model fit, we conclude that the pattern differs across ethnic origin (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Sewasew, Schroeders, Schiefer, Weirich, & Artelt, 2018). Due to the longitudinal nature of the investigation and the procedure of same items for reading self-concept, we permitted for correlations between item residuals (see Marsh et al., 2005). Overlooking this correlated distinctiveness normally effects in positively influenced model fit and biased stability coefficients (Marsh & Hau, 1996). All analyses of REM were conducted twice: with and without controlling for SES. We ran all models by considering some relevant control variables (e.g., longitudinal sampling weight and age of students in months). To cover the second research question (i.e., ethnic origin differences in the mean result across constructs), we made a simple descriptive analysis (mean and standard deviation) and compared the magnitude of ethnic origin disparities using the Cohen's *d* effect size results. We have documented all syntax online at the OSF (Center for Open Science, 2018; https://osf.io/8qsf9/?view_only=ff61d09d3dba49b5b92ae4b22d91e6ac), to make the present analyses transparent and reproducible (Nosek et al., 2015).

3. Results

3.1. Descriptive Statistics

As presented in Table 1, White and Black students have a significantly higher level of reading self-concept than Hispanic and Asian students in grades 3 and 5. However, the effect size of ethnic-background on reading self-concept was small (d =. 01 to. 24**) in grades 3 and 5. Overall, we observed comparable reading self-concept across ethnic-backgrounds. However, White and Asian students significantly surpass Black and Hispanic students in reading competence in grades 3 and 5. There were also moderate to large effects (*d* = .44** to.89**) of the differences in students' ethnic-background on reading competence.

Table 1. Means, standard deviations and effect sizes by ethnic-background for reading self-concept and competence

Time points	Overall	White (a)		Black (b)		Hispanic (c)		Asia (d)		Effect size *(d)*					
	M(SD)	*n*	*M(SD)*	*n*	*M(SD)*	*n*	*M(SD)*	*n*	*M(SD)*	*ab*	*ac*	*ad*	*bc*	*bd*	*cd*
Rsc3	3.22 (.72)	8139	3.24 (0.70)	1881	3.28 (0.76)	2594	3.19 (0.73)	955	3.1 (0.73)	.05	.07	.19**	.12*	.24**	.12*
Rsc5	3.02 (.75)	6471	3.09 (0.72)	1277	3.08 (0.8)	2108	2.90 (0.76)	784	2.91 (0.74)	.01	.23**	.22**	.21**	.20**	.01
Rcom1	.12 (.45)	9420	0.19 (0.42)	2357	-0.06 (0.48)	2580	-0.01 (0.45)	1038	0.22 (0.45)	.55**	.46**	.07	.10	.60**	.51**
Rcom3	.79 (.39)	8094	0.88 (0.28)	1844	0.63 (0.3)	2582	0.67 (0.31)	957	0.81 (0.28)	.86**	.71**	.25**	.13*	.62**	.47**
Rcom5	1.05 (.29)	6465	1.13 (0.27)	1274	0.88 (0.28)	2104	0.94 (0.27)	785	1.06 (.28)	.89**	.70**	.25**	.22**	.64**	.44**

Note. n = 15,395*;* Rsc3 = Reading Self-concept grade 3; Rsc5 = Reading Self-concept grade 5; Rcom1 = Reading Competence grade 1; Rcom3 = Reading Competence grade 3; Rcom5 = Reading Competence grade 5. ** *p* <.01, * *p* <. 05.

The latent correlation matrix between the study variables can be seen in Table 2. Small to moderate positive and significant associations (range *r* =.25** to.51**) were found between reading self-concept and competence. Strong within-domain correlations (range *r* =.70** to.85**) were evidenced among reading competencies across the waves, showing a robust level stability over time; however, a moderate association was revealed between reading self-concept measures, indicating the stability of reading self-concept was still sizable. The mean analyses showed that students' reading competence increased over time, whereas the mean of reading self-concept was stable but slightly decreased over time.

Table 2. Means, standard deviations and latent inter-correlations of reading self-concept and competence

	M	SD	Rsc3	Rsc5	Rcom1	Rcom3	Rcom5

Rsc3	3.22	.72	-				
Rsc5	3.02	.75	.51**	-			
Rcom1	.12	.45	.25**	.33**	-		
Rcom3	.79	.31	.30**	.42**	.77**	-	
Rcom5	1.05	.29	.29**	.44**	.70**	.86**	-
Missingness	-	-	11.8%	19.0%	13.7%	12.4%	18.1%

Note. n = 15,395*;* Rsc3 = Reading Self-concept grade 3; Rsc5 = Reading Self-concept grade 5; Rcom1 = Reading Competence grade 1; Rcom3 = Reading Competence grade 3; Rcom5 = Reading Competence grade 5. ** p <.01, * p <. 05.

3.2. Measurement Invariance Testing Over Time and Ethnic-Background

Prior to the main analyses of the study, we ran longitudinal (MI), and then multi-group confirmatory factor analyses (MGCFA). The longitudinal MI testing procedure consists of a sequence of increasingly restrictive models. The most restrictive form of longitudinal measurement invariance was strict MI, which was supported by the data (i.e., without/with SES: n = 16,178, χ^2 = 603.80, df = 50, CFI = .982; RMSEA = .033, and n = 16,178, χ^2 = 495.03, df = 72, CFI = .991; RMSEA = .026, respectively; for a summary of all steps of longitudinal measurement invariance testing see supplemental material Tables S1 & S2, upper part). Subsequently, the model with constraints of strict longitudinal measurement invariance was taken as a baseline model to test for measurement invariance across ethnic-background groups. We applied a two-index presentation strategy (Hu & Bentler,1999), that combines an absolute fit index such as *Root Mean Square Error of Approximation* (RMSEA), and an incremental fit index such as *Comparative Fit Index* (CFI), as the *chi-square* statistic is not a decisive indicator of model fit. Also, in the case of MI testing across ethnic-background groups, the strict measurement invariant model holds (i.e., without/with SES: n = 16,178, χ^2 = 983.71, df = 230, CFI = .970; RMSEA = .036, and n = 16,178, χ^2 = 822.75, df = 318, CFI = .974; RMSEA = .029, respectively; for a summary of all steps of multi-group longitudinal measurement invariance testing see supplemental material Table S1 & S2, lower part). Accordingly, all subsequent structural equation models were specified with constraints of strict longitudinal MI and strict MI across groups.

3.3. Reciprocal Effects Between Reading Self-Concept and Reading Competence

We designed a two-level analysis procedure following the recommendation of Marsh, et al. (2005). In the first set of analyses, to examine the stability and cross-path effects of individual differences in reading self-concept and competence from grades 1 to 5, we ran cross-lagged models with latent variables in the overall sample (see figure 1, Model 1). The analysis was conducted without and with controlling for the highest SES of the parents at a given time point. Overall, the model fit indices were satisfactory for both models (see the footnotes in Figure 1). There was a noticeable, significant effect from prior reading self-concept on later reading self-concept (β =. 42***) between grades 3 and 5. Likewise, reading competence as measured by standardized test scores in grade 1 significantly affected test scores in grade 3 (β =. 75***), and grade 3 test-score strongly influenced test scores in grade 5 (β =. 84***). That is, a robust autoregressive effect was obtained more for reading test scores than self-concept. In other words, inter-individual differences in reading test scores are more stable or continue over time more than does reading self-concept (see Figure 1). All of these results were also replicated when we considered SES of the family in the model (see details in Figure 1).

The bidirectional relations were revealed between reading self-concept and competence from the paths of grades 3 to 5; that is, early reading self-concept positively and significantly predicted the individual difference in later reading competence (β =. 08***) over and above the effect of prior reading competence. Also, prior reading competence positively and significantly predicted the individual differences in later reading self-concept (β =. 25/. 30***) beyond the effect of prior reading self-concept at both time periods. To compare the self-enhancement with the skill-development effects, we ran two nested models: the main model with freely estimated regressions vs. a model in which the self-enhancement and the skill development effects are constrained to be equal. The comparison of the two models showed a significant deterioration in model fit (CFI ≥ .01) and associated χ^2-difference ($\Delta\chi^2$ (1, N =16,178) = 5.49, p <.001). Hence, the comparison of the reciprocal relations between reading competence and self-concept more strongly supported the skill development part within the REM model, that is larger effects of reading competence on self-concept were found than vice versa (see Figure 1). A comparable pattern of results was also present when we take into account the SES of the family in the model (see details in Figure 1). Because of missing a baseline measure of reading self-concept in grade 1, we only tested the skill development effects from reading competence in grade 1 to reading self-concept in grade 3. Nevertheless, to examine the effects of earlier reading competence on later reading self-concept in the main model, we estimated another REM for grades 3 and 5 only (see supplementary analyses for more details, Figure S1). Generally, the results mirror the ones of the main model reported here.

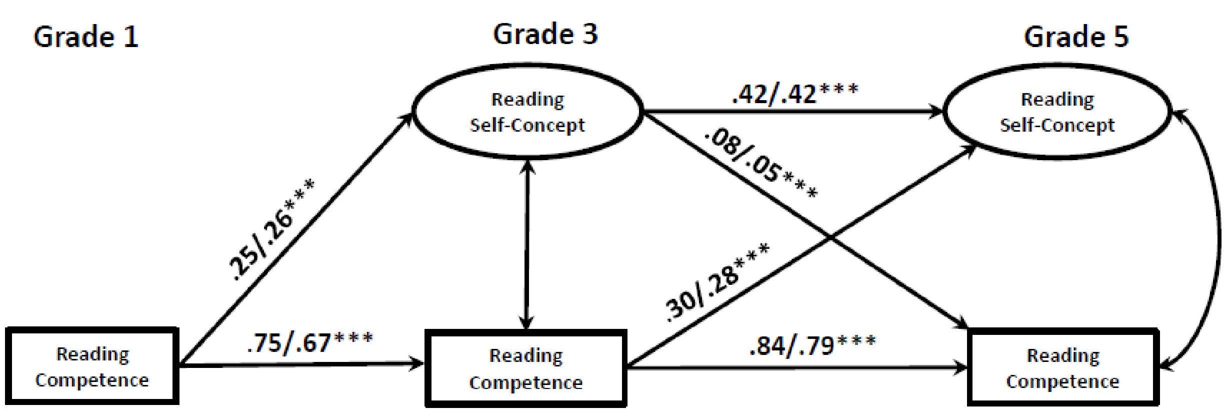


Figure 1. Results of the reciprocal effects model using standardized path coefficients (Model 1). *Note*. The parameters on the left side of the slash are without controlling SES, on the right controlled for SES. The significance asterisk* functions for both regression coefficients. The fit indices of the models without SES are (n = 16,178, χ^2 = 603.80, df = 50, CFI = .982; RMSEA = .033) and with SES (n = 16,178, χ^2 = 495.03, df = 72, CFI = .991; RMSEA = .026). Correlated uniqueness is considered across time points (not depicted); n = 16,178, *** p <.001.

3.4. Reciprocal Relations Between Reading Self-Concept and Reading Competence Across Ethnic-Backgrounds

In the second set of analyses, we executed multi-group REM analyses between reading self-concept and competence in grades 1 to 5, and the model showed a satisfactory fit (see the footnote in Figure 2). The stability results were varied: comparable and strong inter-individual differences of stability coefficients were observed from early to later reading competence between grades 1 and 5 across ethnic-backgrounds; however, compared to test scores, weaker and more varied stability was obtained from prior to later reading self-concept across ethnic-origin, with the highest stability being for Whites and the lowest stability for the Black students' group. Comparable patterns of these results were also evidenced when we included SES of the family in the model (see details in Figure 2).

Overall, the pattern of the causal cross-path model did not show the reciprocal relations between reading self-concept and competence for grades 3 and 5 (except for White students). However, some significant and strong unidirectional path effects (i.e., skill development) were shown from reading competence to self-concept along all paths from grade 1 to 5 for all ethnic-backgrounds (see Figure 2). We also conducted further analyses to test significances for possible ethnic-background differences in the pattern of reciprocal relations using an ethnic-invariant model. For this purpose, we ran two models: First, the factor loadings, intercepts, and residuals were set to be equal, but the regression coefficients were allowed to vary across ethnic origins (see supplemental material, Table S1 & S2). In a second model (unreported supplementary analyses), in addition to the measurement parameter constraints, the regressions were constrained to be equal across ethnic groups. To determine whether the changes in the two models were significant or not, we used a model fit statistics (CFI ≥ .01) and chi-square difference test for the reading standardized test model ($\Delta\chi^2$ (48, *N* =11,788) = 211.97, *p* <. 001). Thus, the reciprocal relations between reading self-concept and competence did vary across the ethnic origins of students; however, note that the REM was observed for White students only. All of these results were also evident when we considered SES of the family in the model (see details in Figure 2).

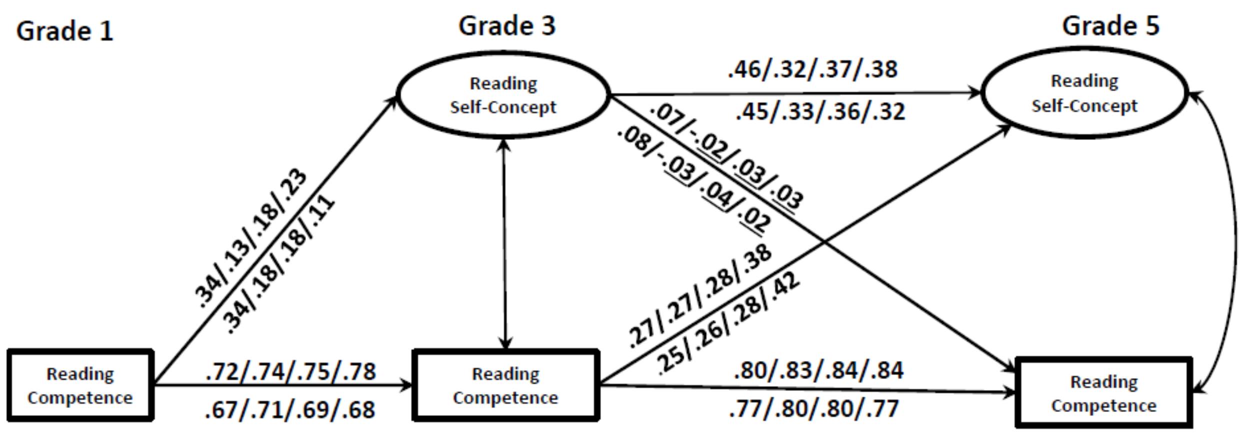


Figure 2. Results of the reciprocal effects model across ethnic-background using standardized path coefficients (Model 2). *Note*. The parameters of on the upper side (White/Black/Hispanic/Asian) of the path line are without controlling SES, on the lower side controlled for SES. All path coefficients were significant (*p* <.001) except the underlined ones. The fit indices of the models are without SES (*n* = 16,178, χ^2 = 983.71, *df* = 230, CFI = .970; RMSEA = .036) and with SES (*n* = 16,178, χ^2 = 822.75, *df* = 318, CFI = .974; RMSEA = .029). Correlated uniqueness is considered across time points (not depicted); *n* = 16,178, *** *p* <.001.

Footnote (result section)

[2] Previous studies suggested that academic achievement is impacted by socialization processes through the influence of social elements on students' competence beliefs, interest or value and emotional engagement with a subject (Eccles, Wigfield, Harold, & Blumenfeld, 1991; Wigfield & Eccles, 2002). The present study examined whether there were variant relations between reading competence and reading self-concept by sex of student. Thus, we estimated models with the relation between reading competence and reading self-concept controlled to be equal between each of the two groups with models in which this relation was allowed to vary freely between the groups. No considerable decline in the model fit statistics was verified; this supports the postulation of a comparable association for boys and girls. The model fit for sex was $\chi^2(112) = 711.09$, CFI=0.98, TLI=0.987, RMSEA=0.033 for the model with parameters varying freely and $\chi^2(128) = 816.69$, CFI=0.978, TLI=0.975, RMSEA=0.033 for the model with parameters constrained to be equal. In addition to the bivariate analyses of REM across sex groups, we also examined the latent mean structure by comparing the construct means on a latent level in a correlated factor model. Latent means for male students were fixed to zero and freely estimated for the female group; that is, positive values indicated a female advantage. For the correlational model, females scored a significantly higher reading self-concept (M = 0.16/0.41), whereas the sex differences for reading competence were not considerable except in grade 1 (M = 1.31/0.06/0.02). This means that females perceived themselves on average as being more competent in reading, although this self-evaluation did not correspond to higher reading competence scores.

4. Discussion

The primary goal of this study was to examine the reciprocal relations between reading self-concept and competence among primary school-aged children from grades 1 to 5. In addition, we investigated the moderating role of students' ethnic origin in the reciprocal relations of reading self-concept and competence. We hypothesized that reading self-concept and competence would be *mutually related* for primary school-aged children; however, the developmental relation was not expected to be *moderated* by ethnic origins of students.

4.1. Reciprocal Effects Between Reading Self-Concept and Reading Competence

Aligned to our first hypothesis, results show the full reciprocal effect (with and without SES) between reading self-concept and competence among primary school-aged children in the overall model (Model 1). That means that reading self-concept and competence are developmentally interlocked—reciprocally related and mutually reinforcing over time. When the existing literature in the area is examined, this result is one the few studies (e.g., Guay et al., 2003) to provide full support for REM in a primary education sample using comprehensive subject-specific standardized test scores. However, our study is different from theirs in that we used domain-specific self-concept and standardized test scores in a large representative U.S. sample with/without consideration of SES of the family. However, the study by Guay and colleagues was confined to domain-general achievement measures (i.e., cumulative teachers' ratings in reading, writing, and math) and perceived competence beliefs of the small French student sample. Our study also different from Walgermo et al.'s (2018) study, as they used only a grade 1 sample of non-normal students (children at risk of reading difficulties) to investigate the effect of early reading interventions. Overall, a strong and significant part of the skill development effect of reading competence on self-concept was evidenced. That means there is path predominance between the relation of reading competence and self-concept in primary school.

4.2. The Moderational Role of Ethnic-Background in the Developmental Relations of Reading Self-Concept and Reading Competence

We also tested whether there are any ethnic-related background differences in the reciprocal relations between variables (i.e., reading self-concept and competence). In contrast to our second hypothesis, we found strong support for the non-generalizability of REM across ethnic-background groups; stated differently, ethnic-background is a potential moderator of the reciprocal effects in primary school-aged children. The result is consistent with recent studies of reading competence and interest in the German secondary school sample (Miyamoto et al., 2017; Schaffner et al., 2016), with the reciprocal relations being evidenced only for native secondary school students. However, our finding is contrary to another analysis of Germany secondary school samples (Möller et al., 2014) which showed that the reciprocal relations between verbal achievement and self-concept are not restricted to whether students have an immigration background or not.

Although Asian ethnic-minority students showed better performance among other minority ethnicities in reading competence, comparable to White-majority students (Peng & Wright, 1994), the positive result of Asian minority students' reading competence in the present analysis vanished when considering reading self-concept in the REM; this signifies that reading self-concept contributes to a share of the variance in the robust competence change among Asian minority students. One explanation for the partial evidence of REM only for White-majority students is the home situation, particularly language barriers. For example, speaking a different language in the home (August et al., 2009; Hall, 2014) and associated inconsistent support for children in reading homework activities, might account for the absence of strong relations between reading competence and self-concept among minority students. Another possibility is that White-majority students have a stronger vocabulary knowledge than students of ethnic-minority students (Powell et al., 2012). If true, this might facilitate their reading competence development and positive perceived competence, eventually resulting in positive bidirectional relations between these variables. This can also be partly explained by the stereotype threat theory, whereby the negative stereotype of a group influences their members' performance (Steele & Aronson, 1995).

4.3. Limitations and Implications

The present study has definite limitations. Firstly, although temporal ordering of study variables is defined by design, it is not possible to make causal assertions on the basis of longitudinal studies alone. In the present context, we conceptualized the term “effect” in its orthodox statistical sense and customary path analytical language, that is, indicating a relation that is not necessarily causal (see also Marsh et al., 2017). Hence, we imply the associations behind the longitudinal interplay of reading self-concept and competence. Nevertheless, more intervention investigations are needed to test causal REM effects in primary school settings.

Secondly, we employed standardized test scores as reading competence measures. In REM, there is a difference between test scores and grades in relation to self-concept. For example, meta-analytic studies report that the impact of competence self-concept is stronger for grades than for test scores (see Huang, 2011; Valentine et al., 2004), because students have no opportunity and little incentive to study for low-stake tests (Möller et al., 2014). Furthermore, low-stake tests have no immediate feedback or direct implications for students’ present or future educational goals (Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006). Therefore, it is noteworthy to combine both achievement measures (i.e., test scores and grades) and self-concept to give a complete picture of their developmental interplay in the elementary school period—where studies are still scarce.

Thirdly, in our study, reading self-concept in grade 1 was not measured, as students were not mature enough to read the text and understand self-concept measures; as a result, reading self-concept measures were omitted from the test administration at this age. Hence, more investigations might consider reading self-concept measures in the first grade of primary school, either using different options of eliciting a response (e.g., happy/sad faces; Davies & Brember, 1994) from young children, or changing the question format (e.g., "Are you a good reader?") rather than the usual declarative format (e.g., "I am a good reader.") to reduce the linguistic complexity faced by young children (Chapman & Tunmer, 1995).

Lastly, when contending theoretical models are tested utilizing secondary data, it is essential to remark on the inadequacy of the items. Researchers can gather data with items in a more scientific and logical manner, which stem from theory-driven research problems. However, secondary data sources might provide partial outcomes as they elicited the data with dissimilar theoretical models in consideration. For example, the item "I think I'm good at reading" could be either "I think I'm good at reading compared to my classmates" if the focus is on social comparison processes, or "I think I'm good at reading compared to my previous achievements" if the focus is on the developmental aspect (Sewasew & Schroeders, in press). We realize that item phrasing can noticeably alter the results, which is a piece that has not been investigated thoroughly. In the current investigation, the reading self-concept measures are articulated specifically; that might also an avenue of future study, expanding their use in many areas of competence investigations and age brackets. Moreover, the self-rating of ability among culturally different populations should be considered, as students of differing family cultural values might have strikingly differing perception of self-competence (Byrne, 1996). For instance, people from Asian countries have found to downplay their self-perception of competence, whilst the reverse is true in countries such as the US (Markus & Kitayama, 1991).

The magnitude and direction of association between academic self-concept and achievement have important implications for parents, educators, and educational and developmental psychologists who work in school settings (Marsh et al., 2005). Given that our findings verified the reciprocal effects model in primary school, teachers and parents should aim to improve reading self-concept and competence simultaneously (Hübner, Wille, Cambria, & Trautwein, 2017). However, we have to be cautious about accepting an "either-or" approach—*either* prior achievement leads to subsequent academic self-concept (a skill development effect) *or* prior academic self-concept leads to subsequent achievement (a self-enhancement effect). For instance, as stated by Marsh and Craven (2006): "If practitioners enhance self-concepts without improving performance, then the gains in self-concept are likely to be short-lived…. If practitioners improve performance without also fostering participants' self-beliefs in their capabilities, then the performance gains are also unlikely to be long-lasting" (p.159). As a result, parental support, reading lessons and reading programs should endeavor to integrate the enhancement of reading self-concepts with that of reading skills. Most importantly, given that we found a differential effect of ethnicity in the REM between reading self-concept and competence, special attention should be given to non-White groups to enrich their reading self-concept and performance status concurrently. One strategy which involves a way of fostering students' reading competence and reading self-concept is CORI (concept-oriented reading instruction, e.g., Guthrie et al., 2004). The CORI rests on the assumption that motivation (reading self-concept and interest), engagement (reading involvement) and development of reading comprehension are strongly influenced by the instructional practices directed towards them. For example, when students are encouraged by interacting with the text which is simultaneously motivating instruction, students' engagement increases and this results in higher reading comprehension. In the same vein, instructional activities synchronized with engagement in reading leads students to become involved and invest efforts, eventually resulting in good performance in reading comprehension. The tasks might include regular reading instruction in the classroom and at home, eventually helping students to grow into competent and self-confident readers.

The divergent outcomes of ethnic-background in relation to self-concept, as compared to reading achievement, provide evidence for the determinant effects of ethnic-background in patterning academic outcomes. Consequently, efforts should be made to narrow achievement disparities, although this cannot fully succeed as long as social stratification caused by ethnic origin is unresolved (Bécares & Priest, 2015). Intervention solutions should also be considered when designing efforts to close achievement gaps; some argue that performance disparities are a product of a more general cognitive process and may be more amenable to change than previously thought (Good, Aronson, & Inzlicht, 2003). However, recently scholars are strongly arguing that for the academic disparity based on ethnic-background, social stratification effects caused by poverty should not be a reason for the poor reading competence, as many students at the wealthier schools and from educated families do not perform very well either. Instead, if students are taught the right way (i.e., learning to read) their competence should be improved as well (Hanford, 2018). In the conventional curriculum, the traditional philosophy of balanced literacy— provide children a lot of good reading books, and through little direction and sufficient exercise, they will become good readers – is not sustainable enough to improve their performance in reading. Hence, teachers need to be taught the science of learning to read (such as how to connect sounds with letters—phonics) and avoid their deeply-held beliefs that reading is a natural development, similar to learning to talk. Simultaneously, explicit and systematic phonics instruction is crucial for students, to teach them about learning to read well (Hanford, 2018).

Finally, it is worth noting that our control variable, SES, is conceivably a valuable indicator and predictor of students' reading competence self-concept level, i.e., categorizing their academic competence and self-concept. For example, meta-analyses showed that SES has a moderate effect on academic competence (see Hattie, 2008). In the current study, we considered SES as one of the essential home-related variables, used it as a controlling covariate, and revealed that it had an impact on REM and ethnic grouping. Consequently, it is reasonable to place more emphasis on the reading competence development of children from lower SES families (Jung, 2014). This is particularly relevant in the U.S. as SES inequalities are channeled by racial prejudice and discriminations at structural and individual levels, thus exerting a solid impact on academic and nonacademic disparities (Bécares & Priest, 2015). It is noteworthy that the relation between SES and academic achievement could be moderated by methodological nature (such as the type of SES measure, and the source of SES data), or by student nature (such as grade level, minority status, and school location). For example, tools of SES that join two or more measures have stronger associations than any solo indicator, and home environment measures have stronger associations than ensured by any unaccompanied or joint group of usual SES indicators (e.g., income and parent education; White, 1982). Besides, when the SES is assessed as a continuous variable, taken from secondary sources and parents instead of students, or gauged from older students, the effect sizes are larger (Sirin, 2005). Moreover, minorities are more likely to live in low-income households, single-parent families, have less educated parents and attend under-funded schools. All of these factors are components of SES and linked to reading competence.

4.4. Final Remarks

In conclusion, the present study confirmed that reading self-concept and reading competence are developmentally interlocked—reciprocally related and mutually reinforcing in primary school-aged children. However, the effect of prior reading competence on reading self-concept was stronger than the reverse. The study highlights strong support for the non-generalizability of REM across ethnic-background groups. That is, ethnic-background was a potential moderator of the reciprocal effects in primary school-aged children, even after partialling out SES. Based on the descriptive analyses, ethnic-background can have detrimental effects on patterning academic outcomes, in contrast to the effects on reading self-concept. However, a large effect size of ethnic imbalance—particularly for White and Asian over Black and Hispanic origin students – was found in reading competence. The study stresses the importance of attaining ethnic achievement parity through eliminating social stratification effects and including intervention-based schemes (e.g., learning to read rather than reading to learn) to curve reading deficits at early stages of schooling.

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

August, D., Shanahan, T., & Escamilla, K. (2009). English language learners: Developing literacy in second-language learners—Report of the national literacy panel on language-minority children and youth. *Journal of Literacy Research*, *41*(4), 432–452. https://doi.org/10.1080/10862960903340165

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among eighth-grade students: Findings from an intersectionality approach. *Public Library of Science One*, *10*(10), e0141363. https://doi.org/10.1371/journal.pone.0141363

Byrne, B. M. (1996). *Measuring Self-concept Across the Lifespan. Washington: American Psychological Association*.

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*(2), 136–145. https://doi.org/10.1037/0022-0663.69.2.136

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *The British Journal of Educational Psychology*, *67* (Pt 3), 279–291; discussion 339-343.

Chapman, James W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, *87*(1), 154–167. https://doi.org/10.1037/0022-0663.87.1.154

Chapman, James W., Tunmer, W. E., & Prochnow, J. E. (2000). Early reading-related skills and performance, reading self-concept, and the development of academic self-concept: A longitudinal study. *Journal of Educational Psychology*, *92*(4), 703–708. https://doi.org/10.1037/0022-0663.92.4.703

Chen, S.K., Yeh, Y.C., Hwang, F.M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. *Learning and Individual Differences*, *23*, 172–178. https://doi.org/10.1016/j.lindif.2012.07.021

Davies, J., & Brember, I. (1994). The reliability and validity of the "smiley" scale. *British Educational Research Journal*, *20*(4), 447–454.

D'hondt, F., Van Praag, L., Van Houtte, M., & Stevens, P. A. (2016). The attitude–achievement paradox in Belgium: An examination of school attitudes of ethnic minority students. *Acta Sociologica*, *59*(3), 215–231. https://doi.org/10.1177/0001699316636944

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, *15*(1), 35–54.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1991). Age and gender differences in children's self- and task perceptions during elementary school, *Child Development*, *64* (3), 830 - 847.

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

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence*, *29*(6), 679–695. https://doi.org/10.1023/A:1026455906512

Fraine, B. D., Damme, J. V., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, *32*(1), 132–150. https://doi.org/10.1016/j.cedpsych.2006.10.005

Geiser, C. (2013). *Data Analysis with Mplus*. New York: Guilford Press.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*(6), 645–662. https://doi.org/10.1016/j.appdev.2003.09.002

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., … Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96*(3), 403–423. https://doi.org/10.1037/0022-0663.96.3.403

Hall, Lily, "Examining the relationship between reading ability and reading self-concept in differing socio-economic schools" (2014). *Open Access Master's Theses.* Paper 296.

http://digitalcommons.uri.edu/theses/296.

Hanford, E. (2018). Why aren't kids being taught to read? Retrieved September 15, 2018, from https://www.apmreports.org/story/2018/09/10/hard-words-why-american-kids-arent-being-taught-to-read.

Hattie, J. (2008). *Visible Learning* (1st ed.). London ; New York: Routledge.

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), 1–55. https://doi.org/10.1080/10705519909540118

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Hübner, N., Wille, E., Cambria, J., & Trautwein, U. (2017). Maximizing gender equality in STEM by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM. *Journal of Educational Psychology,*109, 993-1009. https://doi.org/10.1037 / edu0000183.

Jung, E. (2014). Examining differences in kindergarteners' mathematics learning: A closer look at instruction, socioeconomic status, and race. *The Journal of Educational Research*, *107*(6), 429–439. https://doi.org/10.1080/00220671.2013.833074

Kigel, R. M., McElvany, N., & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation: A longitudinal study. *Learning and Instruction*, *35*, 73–84. https://doi.org/10.1016/j.learninstruc.2014.10.001

Little, T. (2013). *Longitudinal Structural Equation Modeling*. Guilford Press.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 224–253.

Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, *84*(1), 35–42. https://doi.org/10.1037/0022-0663.84.1.35

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal*, *34*(4), 691–720. https://doi.org/10.3102/00028312034004691

Marsh, Herbert W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*(4), 646–656.

Marsh, Herbert 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. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, Herbert W., & Hau, K.T. (1996). Assessing goodness of fit. *The Journal of Experimental Education*, *64*(4), 364–390. https://doi.org/10.1080/00220973.1996.10806604

Marsh, Herbert W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, Herbert W., Pekrun, R., Murayama, K., Arens, K. A., Parker, P. D., Guo, J., & Dicke, T. (2017). An integrated model of academic self-concept development: academic self-concept, grades, test scores, and tracking over six years. *Developmental Psychology*, 54, 263–280. http://dx.doi.org/10.1037/dev0000393.

Marsh, Herbert 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*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Mead, G. H. (1934). *Mind, Self, and Society*. Retrieved from https://www.press.uchicago.edu/ucp/books/book/chicago/M/bo20099389.html

Miyamoto, A., Pfost, M., & Artelt, C. (2017). Reciprocal relations between intrinsic reading motivation and reading competence: A comparison between native and immigrant students in Germany: Reading Motivation and Reading Competence. *Journal of Research in Reading*. https://doi.org/10.1111/1467-9817.12113

Möller, J., & Pohlmann, B. (2010). Achievement differences and self-concept differences: Stronger associations for above or below average students? *British Journal of Educational Psychology*, *80*(3), 435–450. https://doi.org/10.1348/000709909X485234

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An Integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 0002831211419649. https://doi.org/10.3102/0002831211419649

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, *84*(4), 591–611. https://doi.org/10.1111/bjep.12047

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children*, *73*(2), 165–183. https://doi.org/10.1177/001440290707300203

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*(4), 1170–1191. https://doi.org/10.1037/a0036307

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

OECD. (2010). PISA 2009 results: Overcoming social background equity in learning opportunities and outcomes (Volume II): Equity in learning opportunities and outcomes. *OECD Publishing*.

Ogbu, J. U. (1993). Differences in cultural frame of reference. *International Journal of Behavioral Development*, *16*(3), 483–506. https://doi.org/10.1177/016502549301600307

Peng, S. S., & Wright, D. (1994). Explanation of academic achievement of asian american students. *The Journal of Educational Research*, *87*(6), 346–352. https://doi.org/10.1080/00220671.1994.9941265

Pollack, J. M., Najarian, M., Rock, D. A., & Atkins-Burnett, S. (2005). Early childhood longitudinal study, kindergarten class of 1998? 99 (ECLS-K). Psychometric report for the fifth grade. NCES 2006? 036. ED Pubs, P. Retrieved from https://eric.ed.gov/?id=ED489079

Powell, D. R., Son, S.-H., File, N., & Froiland, J. M. (2012). Changes in parent involvement across the transition from public school prekindergarten to first grade and children's academic outcomes. *The Elementary School Journal*, *113*(2), 276–300. https://doi.org/10.1086/667726

Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction*, *29*, 21–30. https://doi.org/10.1016/j.learninstruc.2013.07.004

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Rodriguez, P. J. (2014). A review of "minority status, oppositional culture, & schooling." *Educational Studies*, *50*(3), 294–298. https://doi.org/10.1080/00131946.2014.907166

Rosseel, Y. (2012). *lavaan:* An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36. http://www.jstatsoft.org/v48/i02/

Rosenbloom, S. R., & Way, N. (2004). Experiences of discrimination among african american, asian american, and latino adolescents in an urban high school. *Youth & Society*, *35*(4), 420–451. https://doi.org/10.1177/0044118X03261479

Salikutluk, Z. (2016). Why do immigrant students aim high? Explaining the aspiration–achievement paradox of immigrants in Germany. *European Sociological Review*, *32*(5), 581–592. https://doi.org/10.1093/esr/jcw004

Schaffner, E., Philipp, M., & Schiefele, U. (2016). Reciprocal effects between intrinsic reading motivation and reading competence? A cross-lagged panel model for academic track and nonacademic track students. *Journal of Research in Reading*, *39*(1), 19–36. https://doi.org/10.1111/1467-9817.12027

Schiefele, U., Schaffner, E., Moller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, *47*(4), 427–463.

Schroeders, U., & Gnambs, T. (2018). Degrees of freedom in multigroup confirmatory factor analyses. *European Journal of Psychological Assessment*, 1–9. https://doi.org/10.1027/1015-5759/a000500

Schroeders, U., & Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, *71*, 849–869. https://doi.org/10.1177/0013164410391468

Sewasew, D., Schroeders, U., Schiefer, I. M., Weirich, S., & Artelt, C. (2018). Development of sex differences in math achievement, self-concept, and interest from grade 5 to 7. *Contemporary Educational Psychology*, *54*, 55–65.

Sewasew, D., & Schroeders, U. (in press). The developmental interplay of academic self-concept and achievement within and across domains among primary school students. *Contemporary Educational Psychology.*

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Singh, K., Chang, M., & Dika, S. (2010). Ethnicity, self-concept, and school belonging: effects on school engagement. *Educational Research for Policy and Practice*, *9*(3), 159–175. https://doi.org/10.1007/s10671-010-9087-0

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, *67*(2), 135–149. https://doi.org/10.1080/00220979909598349

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles*, *50*(3–4), 241–252.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

Steele, Claude M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613.

Steele, Claude M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology* (Vol. 34, pp. 379–440). Elsevier.

Tourangeau, K., Brick, M., Le, T., Wan, S., Weant, M., Nord, C., … Walston, J. (2004). User's manual for the ecls-k third grade. public-use data file and electronic code book. Early childhood longitudinal study, kindergarten class of 1998-99. NCES 2004-001. ED Pubs, P. Retrieved from https://eric.ed.gov/?id=ED484511

Tourangeau, K., Nord, C., Le, T., Sorongon, A. G., & Najarian, M. (2009). Early childhood longitudinal study, kindergarten class of 1998-99 (ECLS-K): Combined user's manual for the ecls-k eighth-grade and k-8 full sample data files and electronic codebooks. NCES

2009-004. National Center for Education Statistics. Retrieved from https://eric.ed.gov/?id=ED511826

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Walgermo, B. R., Foldnes, N., Uppstad, P. H., & Solheim, O. J. (2018). Developmental dynamics of early reading skill, literacy interest, and readers' self-concept within the first year of formal schooling. *Reading and Writing*, *31*(6), 1379–1399. https://doi.org/10.1007/s11145-018-9843-8

Warmington, M., & Hulme, C. (2012). Phoneme Awareness, Visual-Verbal Paired-Associate Learning, and Rapid Automatized Naming as Predictors of Individual Differences in Reading Ability. *Scientific Studies of Reading*, *16*(1), 45–62. https://doi.org/10.1080/10888438.2010.534832

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, *29*, 39–47. https://doi.org/10.1111/j.1745-3992.2010.00182.x

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *A Vol. in the educational psychology series. Development of achievement motivation* (pp. 91-120). San Diego, CA, US: Academic Press. http://dx.doi.org/10.1016/B978-012750053-9/50006-1

Supplemental materials

Appendix A: Measurement invariance testing across time and ethnic-background

Appendix B: Results of measurement invariance testing over time and ethnic-background

Table S1. The goodness of fit indices for longitudinal and multi-group MI testing without SES

Table S2. The goodness of fit indices for longitudinal and multi-group MI testing with SES

Figure S1: Results of the REM only using grade three and five reading self-concept and reading competence.

Appendix A: Measurement invariance testing across time and ethnic-background

In longitudinal studies, measurement invariance (MI) testing across time points is pivotal to make valid statements about developmental trends. It is necessary to ensure that the measurement does not change over the course of study (Geiser, 2013). Therefore, we applied longitudinal MI testing (Geiser, 2013; Little, 2013) across time points (grades 1, 3, and 5) to examine the comparability of measurement instruments for reading competence and self-concept. In the baseline model, all measurement parameters (i.e., factor loadings, intercepts, and residual variances) were freely estimated across measurement occasions (*configural model*); only the residuals of similarly-worded items were allowed to correlate across time points. Factor means were fixed at the first time point and freely estimated at all other time points. For *metric invariance,* the factor loadings were fixed to be equal across time points. In addition, intercepts are fixed to be equal in *scalar invariance* testing. In the most restrictive form (*strict invariance*), all measurement parameters are equal across time points. In comparison to MI testing across ethnic-background-groups, which is explained in the next paragraph, in longitudinal MI, the residuals of similarly-worded items are allowed to co-vary across measurement occasions.

For invariance testing across ethnic-background groups, we conducted a type of the most restrictive form of longitudinal measurement invariance we could establish — a series of increasingly restrictive models within the framework of multiple-group confirmatory factor analysis (Schroeders & Wilhelm, 2011; Wicherts & van Dolan, 2010). In the least restrictive model of *configural measurement invariance*, all measurement parameters (i.e., factor loadings, intercepts, and residuals) vary freely in each group. In subsequent steps of measurement invariance testing, additional equality restrictions are imposed on specific parameters across ethnic-background groups as follows: *metric invariance* (factor loadings), *strong invariance* (factor loadings and intercepts), and *strict invariance* (factor loadings, intercepts, and residuals). To make this point clear, in longitudinal MI testing, restrictions are introduced across measurement time points; in MI testing, restrictions are introduced across ethnic-origin groups. Because the complete MI testing is comprised of 5 different models in total, we used a two-step testing procedure. To assess whether imposing additional constraints on measurement parameters leads to deterioration in model fit, the differences in the CFI between two consecutive models was evaluated (Vandenberg & Lance, 2000), whereas a deterioration in a model of CFI ≥ .01 is usually considered indicative of a serious deterioration.

Appendix B: Results of measurement invariance testing over time and ethnic-background

We executed two measurement invariances on the basis of reading test score and self-concept without and with SES. First, for the reading test scores and reading self-concept, the longitudinal MI testing consists of a sequence of increasingly restrictive models (see Table 1, upper part). The most restrictive form of longitudinal measurement invariance was strict MI. Subsequently, this strict form of measurement invariance was taken as a baseline model to test for measurement invariance across ethnic-origin groups. In more detail, in the step of *configural* *invariance*, all factor loadings, intercepts, and residuals were freely estimated; only the overall structure was supposed to be identical across groups. It should be noted that although the measurement parameters are allowed to freely vary *across groups*, the restrictions of longitudinal measurement invariance hold; that is, factor loadings, intercepts, and residuals are fixed to equality *across time points* (see Table 1, lower part). According to the cut-off values reported by Hu and Bentler (1999), model fit indices of the configural model were good: n = 16,178, χ^2 = 804.15, df = 214, p < .001; CFI = .983; RMSEA = .033. In the second step, *metric invariance*, only factor loadings were constrained to be equal across ethnic-origin groups; means were fixed to zero. Additionally imposing the equality constraints of metric invariance did not yield a meaningful deterioration in the model fit (n = 16,178, χ^2 = 835.90, df = 218, p < .001; CFI =. 980; RMSEA =.033, ΔCFI = .003, ΔRMSEA = .000). In the third step, *strong invariance* (also labeled *scalar invariance*), factor loadings and intercepts were constrained to be equal across the ethnic-origin groups, and residuals were freely estimated, which did not result in significant changes in the model fit (n = 11,788, χ^2 = 845.38, df = 221, p < .001; ΔCFI = .007, ΔRMSEA = .000). In addition to the constraints of scalar invariance, in the last step of measurement invariance testing (*strict invariance*) the residuals were also fixed to equality across groups. Even this model held (n = 16,178, χ^2 = 983.71, df = 230, p < .001; ΔCFI = .012, ΔRMSEA = .003). Accordingly, all subsequent structural equation models were specified with the constraints of strict longitudinal MI and strict MI across ethnic-origin groups. Second, for the reading test scores and self-concept with SES, we exactly repeated the same procedure and the detailed results were present in the upper and lower part of Table 2 (see longitudinal MI and MGCFA for reading test scores and self-concept with SES).

Table S1. The goodness of fit indices for longitudinal and multi-group MI testing without SES

1. *Longitudinal MI testing with reading self-concept and competence without SES*							
Models (complete sample)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	357.09	45	. 990	–	.984	.026	
Metric invariance	367.54	47	. 990	.000	.984	.026	.000
Strong invariance	392.25	47	. 989	.001	.983	.027	.001
Strict invariance	603.80	50	. 982	.007	.974	.033	.006
2. *Testing MI across ethnic-background groups for the strict longitudinal MI model without SES*							
Models (MGCFA)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	804.15	212	. 983		.973	.033	
Metric invariance	835.90	218	. 980	.003	.973	.033	.000
Strong invariance	845.38	221	. 982	.001	.973	.033	.000
Strict invariance	983.71	230	. 970	.012	.969	.036	.003

Note. *n* = 16,178; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Table S2. The goodness of fit indices for longitudinal and multi-group MI testing with SES

1. *Longitudinal MI testing with reading self-concept and competence with SES*							
Models (complete sample)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	314.14	67	. 993	–	.989	.021	
Metric invariance	321.65	69	. 993	.000	.989	.021	.000
Strong invariance	343.88	69	. 992	.001	.989	.022	.001
Strict invariance	495.03	72	. 991	.001	.990	.026	.004
2. *Testing MI across ethnic-background groups for the strict longitudinal MI model with SES*							
Models (MGCFA)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	683.02	300	. 989		.985	.025	
Metric invariance	706.25	306	. 986	.000	.984	.025	.000
Strong invariance	722.86	309	. 988	.002	.984	.025	.002
Strict invariance	822.75	318	. 974	.013	.972	.029	.004

Note. *n* = 16,178; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation.

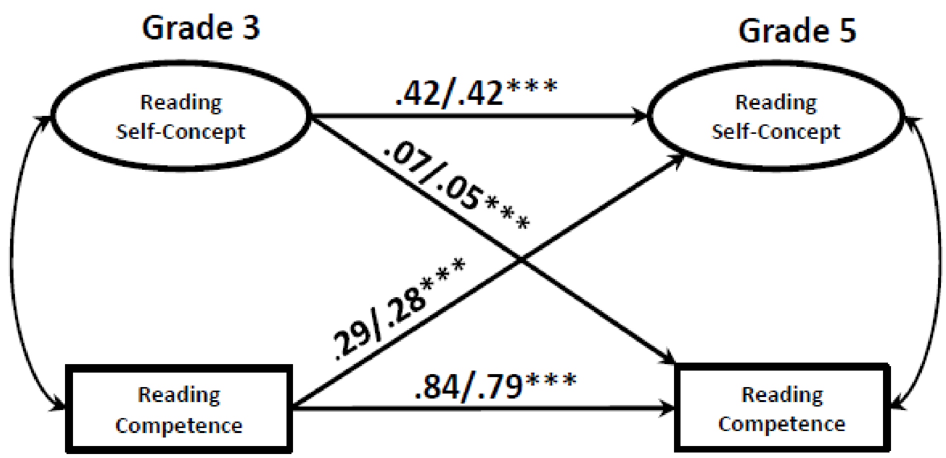


Figure S1. Results of the reciprocal effects model using standardized path coefficients. *Note*. Rcom = Reading competence. The parameters on the left side of the slash are without controlling SES, on the right controlled for SES. The significance asterisk* functions for both regression coefficients. The fit indices of the models without SES are (n = 16,178, χ^2 = 382.89, df = 35, p <.001; CFI = .986; RMSEA =.031) and with SES (n = 16,178, χ^2 = 343.11, df = 47, p <. 001; CFI =. 989; RMSEA =. 027). Correlated uniqueness is considered across time points (not depicted); n = 16,178, *** p <.001.

Development of Sex Differences in Math Achievement,

Self-Concept, and Interest From Grade 5 to 7

Daniel Sewasew[a], Ulrich Schroeders[a, b, d], Irene M. Schiefer[b], Sebastian Weirich[c], Cordula Artelt[a, b, e]

[a] *Bamberg Graduate School of Social Sciences (BAGSS), University of Bamberg, Germany*

[b] *Department of Educational Research, University of Bamberg, Germany*

[c] *Institute for Educational Quality Improvement, Humboldt Universität zu Berlin, Germany*

[d] *Institute of Psychology, University of Kassel, Germany*

[e] *Leibniz Institute for Educational Trajectories (LIfBi), University of Bamberg, Germany*

This paper uses data from the German BiKS project conducted at the University of Bamberg. BiKS is a DFG (German Research Foundation) funded interdisciplinary research project that runs two longitudinal studies on educational processes, competence development, and selection decisions in preschool and school-age children. The secondary school sample of the BiKS-8-14, from 2005–2014 was used

for this study http://doi.org/10.5159/IQB_BIKS_8_14_v2.

Correspondence concerning this article should be addressed to Daniel Sewasew, Bamberg Graduate School of Social Sciences, University of Bamberg, Feldkirchenstraße 21, 96045 Bamberg, Germany, E-mail: sewasewdaniel@gmail.com (daniel.sewasew@uni-bamberg.de), Phone: +49(0)951/863-3617.

Abstract

Sex differences in mathematics achievement have been a controversial topic in educational psychology for a long time. This study sheds light on the developmental aspects of sex differences in math achievement and domain-specific motivational variables such as self-concept and interest. Using a *Reciprocal Effects Model* (REM), we analyzed 2,342 German fifth to seventh grade students who participated in a large-scale longitudinal study. Math self-concept was validated as a consistent predictor of subsequent achievement and interest for both sexes, supporting the *self-enhancement* part of the REM using test-scores and teacher-assigned grades. However, math achievement affects subsequent self-concept inconsistently (i.e., the *skill development* part). Although the bivariate relationships between the constructs were homogeneous across sex and over time, there were large sex differences in the motivational constructs, but not in the achievement measure regardless of achievement measures. The present findings underline the importance of considering both the mean and the covariance structure when describing sex differences in academic achievement. In addition, they also stress the impact of motivational constructs on educational achievement, which also have implications for sex-specific intervention programs in general.

Keywords: sex differences, math achievement, math self-concept, math interest, reciprocal effect model

Development of Sex Differences in Math Achievement,

Self-Concept, and Interest From Grade 5 to 7

1. Introduction

Sex differences in students' academic achievement have been a controversial topic among psychologists and educators for several decades (Halpern et al., 2007; Hyde, 2005). The study of sex differences in cognitive abilities has been seen as "a sociopolitical minefield, with serious implication for a wide range of public policies (e.g., affirmative action, compensatory education, pay equity)" (Halpern & LaMay, 2000, p.230). Although the sex ratio among freshmen in Western universities is at least equal, women are often underrepresented in STEM study programs (science, technology, engineering, and mathematics) - in particular engineering, physical science, and computational science (Hyde, 2014). Small sex differences in math achievement in favor of boys have frequently been reported in secondary education (Brunner et al., 2013), but they are not the major factor explaining differences in the enrollment for STEM subjects. In this context, researchers have emphasized the importance of sex-related differences in motivational constructs such as domain-specific self-concept and interest in course selection and educational choices (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). Given that STEM subjects have been recognized as a critical filter to a higher salary and esteemed jobs, sex differences in STEM abilities or motivation at any age or educational level are of particular interest. Moreover, to tap the large talent pool of female students and to increase equity between sexes regarding participation in STEM studies, it is important to understand more thoroughly the developmental interplay between educational achievement and motivational constructs across sex groups.

In the literature on the interplay of academic achievement and motivational constructs such as domain-specific self-concept and interest, there are three competing perspectives: The *skill development theory* (i.e., academic achievement as a major determinant of academic self-concept), the *self-enhancement theory* (i.e., self-concept as a primary determinant of academic achievement, Calsyn & Kenny, 1977), and the notion of *reciprocal relations* between academic achievement and self-concept which are mutually reinforcing (e.g., Marsh, 1990). The long-standing debate between proponents of skill development vs. self-enhancement has only superficially been settled with the introduction of the reciprocal effects model (Marsh & Craven, 2006) because the answer to the imminent and prevailing question which path is more prevalent is still pending. Consequently, reconsidering the prospect of causal order between math achievement and self-concept/interest has intriguing theoretical and practical implications.

The contribution of the present study is, therefore, threefold: First, we investigated the interplay between math achievement and self-concept and interest, respectively. That is, we study whether academic self-concept and interest have substantial effects on subsequent academic achievement (self-enhancement) or math achievement on the motivational constructs (skill-development). Second, sex-related comparisons in the REM context are only valid if the used measures are invariant over time and across sex groups (Cheung & Rensvold, 2002; Little, Preacher, Selig, & Card, 2007). In this regard, our study also likes to draw attention to the psychometric properties of the measure and an appropriate way of modeling longitudinal data. Third, we added to previous findings by studying the relations among math achievement (i.e., test-scores and grades) and motivational constructs (i.e., self-concept and interest) in a large sample of German secondary school students. In this respect, one aim of the present study is to provide further evidence for the reported literature findings, since “replication evidence is the gold standard by which scientific claims are evaluated" (p. 410, Bonett, 2012). In contrast, unfortunately, only a small number of the publications in psychology are replications (see also Makel, Plucker, & Hegarty, 2012).

1.1. Reciprocal Relations Between Math Achievement and Self-Concept

Math is a core subject in every school program, and success in math is strongly associated with math self-concept, which is defined as the subjective feeling and belief in one's competence in math (Marsh, 1989). In the debate on the causal order of motivational and cognitive constructs, two opposing perspectives were irreconcilable for a long time (Calsyn & Kenny, 1977). On the one hand, the *self-enhancement hypothesis* postulates that self-concept positively affects subsequent achievement via educational decisions, higher motivation, effort, and investment (Marsh & Yeung, 1997). According to this perspective, encouraging academic self-concept should stimulate students' achievement. On the other hand, the *skill-development hypothesis* postulates that achievement positively affects the academic self-concept, which is explained by the fact that students' evaluate their competence (i.e., self-concept) by comparing their performance/grades to others (i.e., social comparison, Möller & Pohlmann, 2010). If a student gets good grades and excels, he or she develops a positive self-concept, whereas students with low achievement develop a low academic self-concept. A contemporary perspective on this motivation-achievement debate is the *Reciprocal Effects Model* (REM), which integrates both ideas by assuming a mutually dependent positive relationship between academic self-concept and achievement over time (Marsh, 1990; Marsh & Craven, 2006). Differently worded, students who are competent in math develop a positive self-concept, and this self-concept, in turn, promotes students' academic performance.

A meta-analysis summarizing the empirical evidence of the *self-enhancement* hypothesis showed a weak to moderate positive relation from academic self-concept to later achievement within a given domain (Valentine, Dubois, & Cooper, 2004; see also Huang, 2011). Similarly, a meta-analysis of the *skill development* model revealed a strong positive path from academic achievement to self-concept (Möller, Pohlmann, Köller, & Marsh, 2009). Moreover, several longitudinal comparable *reciprocal relations* have been found between math achievement and self-concept (Möller, Retelsdorf, Köller, & Marsh, 2011; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014). Nonetheless, findings of the causal relationship between academic self-concept and achievement are inconsistent (Marsh & Martin, 2011) and may be moderated by the school subject and the time point in the course of education. For example, Chen, Yeh, Hwang, and Lin (2013) found reciprocal effects for Taiwanese high-school students in math, but for the language domain, only the skill development part was supported. Moreover, the skill development model only holds for younger children, whereas the self-enhancement effects were present only for older students, revealing a developmental change in the reciprocal effects (Chen et al., 2013; Fraine, Damme, & Onghena, 2007). However, findings of other studies showed that the effect of academic self-concept on achievement was unrelated to age and the school type (Guay, Marsh, & Boivin, 2003; Valentine & DuBois, 2005). Thus, currently, there is insufficient evidence to conclude that the REM varies with age (Marsh, Byrne, & Yeung, 1999; Valentine & DuBois, 2005).

Other explanations for these diverging results might be: a) different operationalization of math achievement (grades vs. test-scores), b) the breadth of the definition of math self-concept, for example, math enjoyment and competence beliefs (Else-Quest, Hyde, & Linn, 2010; Pinxten et al., 2014), as well as math confidence (Else-Quest et al., 2010; Ganley & Lubienski, 2016) as part of self-concept, c) different methodological approaches, that is linear regression with manifest indicators vs. latent variable modeling (Marsh et al., 2005; von Maurice, Dörfler, & Artelt, 2014), d) the average ability level of the sample (e.g., academic track only, see also Marsh et al., 2017), and e) design of the study, that is, cross sectional vs. longitudinal (Else-Quest et al., 2010; Marsh et al., 2005).In the present study, we investigated the causal relationship of math self-concept and achievement across sex groups with standardized achievement test-score and teacher-assigned grade at a latent level using longitudinal data in a heterogeneous sample of secondary school students.

1.2. Reciprocal Relations Between Math Achievement and Interest

There are various conceptualizations of interest (Hidi & Renninger, 2006). In line with the person-object concept of Krapp (2002), we define interest as a characteristic of a relation between a student and a domain. For example, students with high interest are characterized by a consistently high cognitive commitment and emotional involvement in math. In this sense, interest is a) considered domain-specific, b) comprised of a cognitive and an affective component, c) built and nourished over the school career, and d) relatively stable over a variety of situations and over time (Jansen, Lüdtke, & Schroeders, 2016; Krapp, 2002; Marsh et al., 2005). The development of math interest and achievement is intertwined. Math interest is at the central of self-determined activities and likely to be reciprocally related with achievement (Garon-Carrier et al., 2016; Marsh et al., 2005; Ryan & Deci, 2000). That is, students interested in math will be more persistent and keen in pursuing math-related activities and, thus, more likely achieve higher grades/test-scores. Likewise, better grades in math should encourage and motivate students (Ryan & Deci, 2000).

The current literature on REM, including interest as a motivational construct, is scarce and often yields contradictory results: Some researchers (e.g., Liu, 2009; Pinxten et al., 2014; Yoon, Eccles & Wigfield, 1996) found a positive mutual relation between math interest and achievement, while others found a significant skill development effect, but no self-enhancement effect (Ganley & Lubienski, 2016). In contrast, Marsh and colleagues (2005) reported little or no effect of academic achievement on interest, but a significant positive self-enhancement effect. Whereas, Köller and colleagues (2001) argued that early math interest might yield an indirect effect on course selection, which in turn affects math achievement. This notion is also in line with research emphasizing the importance of interest for course enrollment (in comparison to prior achievement), both in secondary and tertiary education (Wigfield, 1994; Wigfield & Eccles, 2000). Overall, the aforementioned empirical inconsistencies raise the question as to what extent students' interest is affected by prior achievement and *vice versa*. Because longitudinal studies with a sufficiently large and heterogeneous sample are rare, the present study tries to complement previous research on this topic.

1.3. Reciprocal Relations Among Math Competence, Self-Concept, and Interest

To our knowledge, the REM was used to examine math achievement, domain-specific self-concept, and interest in concert in only three articles (Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014). The three studies were concordant with respect to the mutual relations between self-concept and achievement. However, even if the effects are bidirectional, the question remains which effect is predominant (i.e., the size of the causal effects). For instance, a fairly balanced reciprocal effect between self-concept and achievement was found (Pinxten et al., 2014), but also a stronger skill developmental paths (Ganley & Lubienski, 2016) as well as a stronger self-enhancement path (Marsh et al., 2005). In summary, these studies provided inconclusive results by demonstrating either balanced reciprocal effects or a stronger skill development effect or stronger self-enhancement effects. Furthermore, results also have been inconsistent regarding the relationship between interest and achievement. Some authors (e.g., Yoon et al., 1996) found support for reciprocal causal relations between intrinsic value and math performance, whereas others revealed no mutual relations between interest and achievement at all (e.g., Marsh et al., 2005). Consequently, the existing accounts fail to resolve the contradictions among the interplay of math performance, self-concept, and interest.

1.4. Reciprocal Relations and Achievement Outcome

Hitherto, in the REM literature students' achievement has been operationalized either with standardized test-scores (Ganley & Lubienski, 2016; Pinxten et al., 2014) or—more often—with teacher-assigned grades (Möller et al., 2011; Niepel, Brunner, & Preckel, 2014). It is important to note that the size of the reported reciprocal relations between motivation and achievement vary depending on the achievement outcome. The common finding is that the relations are stronger for grades than for test-scores (e.g., Marsh, Hau, & Kong, 2002; Marsh & Yeung, 1998; Möller et al., 2014), which might be attributed to the fact that students usually do not get any form of feedback on a standardized test. Thus, there is no salient source of feedback which could affect subsequent school performance. However, in contrast to the aforementioned empirical evidence, Valentine, DuBois, and Cooper (2004) found almost no differences between the different achievement outcomes and self-concept in their comprehensive meta-analysis. Only a few studies have considered both test-scores and grades as indicators of academic achievement and subsequently separately examined their relational interplay (Marsh, 1990; Marsh et al., 2005; Pinxten et al., 2010; Trautwein et al., 2006) or combined them in a latent construct (Marsh & Yeung, 1997; Pinxten et al., 2010; Skaalvik & Valås, 1999). With the exception of Marsh's

(2005) and the present study, the REM including subject-specific self-concept and interest as motivational constructs hasn't been validated for two math achievement measures (i.e., test-scores and grades).

1.5. Sex Differences in the Reciprocal Relations and Latent Means

Testing the generalizability of REM over sex, especially in sex-stereotypic subjects, for instance, math and verbal can help to understand the reasons for a gender gap in STEM subjects (Halpern et al., 2007). Marsh theorized and tested a *differential socialization hypothesis* in which "sex-linked differences in socialization patterns may fail to reinforce adequately boys' positive attitudes, expectations, and performance in verbal areas as well as fail to reinforce adequately girls' positive attitude, expectations, self-concepts, and performance in mathematics" (Marsh, 1989, p. 195). Differential reinforcement directly refers to differential relations in the REM. Based on Marsh's idea, it is expected that the link between math self-concept and achievement could be stronger for boys than for girls. Based on this idea, it is expected that the link between math self-concept and achievement should be stronger for boys than for girls. However, only a few studies have examined this sex-specific assumption. In a longitudinal study among students attending Grade 6 and 7, reciprocal causal relations among math self-concept, interest, and academic achievement were somewhat moderated by the sex and the schooling context (Yoon et al., 1996), that is, math self-concept effect positively affected achievement for males, whereas the relation was negative for females. More recent studies also reported small sex differences (Pinxten et al., 2014), and the strength of effects from math performance to interest may be slightly stronger for girls than for boys (Ganley & Lubienski, 2016). However, on a meta-analytical level, no sex differences were found in the reciprocal relations (Huang, 2011; Valentine, DuBois, & Cooper, 2004). A similar sex-invariant outcome was reported when academic achievement, self-concept, and interest were juxtaposed (see Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014).

With respect to the mean structure, a plethora of empirical studies have documented small to negligible sex differences ($d < 0.20$) in math competence, although the heterogeneity of the results across nations is large ($-0.41 \leq d \leq 0.40$) (e.g., Else-Quest, Hyde, & Linn, 2010; Hyde, Fennema, & Lamon, 1990; Marsh et al., 2005). In general, effect sizes of sex differences in math achievement are in line with the *gender similarities hypothesis*—both sexes are similar on most, but not all psychological variables (Hyde, 2014), irrespective of the psychometric model applied (Brunner et al., 2013). However, such broad generalizations divert the attention from several moderators that might affect the strength of potential sex differences: a) the specific content (Hyde, 2014); b) the complexity of the test (Gibbs, 2010); c) the age of students (Ganley & Lubienski, 2016; Hyde et al., 1990); d) the educational context or country (Brunner et al., 2013; Rustemeyer & Fischer, 2005); and e) (most importantly) how math achievement was measured (Egorova & Chertkova, 2016). For instance, girls had higher math achievement based on school grades, but a lower achievement based on test scores (Egorova & Chertkova, 2016). Moreover, boys outperformed girls when math tests contained geometrical material and less algebra (Hyde, 2014). There is also evidence that girls performed slightly better in elementary school, whereas boys outperform girls in high school and college years (Hyde et al., 1990).

In contrast to achievement, substantial sex differences have been reported for math self-concept and interest. In general, boys had higher levels of math self-concept (Else-Quest et al., 2010; Pinxten et al., 2014) and interest (Frenzel, Pekrun, & Goetz, 2007; Ganley & Lubienski, 2016; Pinxten et al., 2014). From a developmental perspective, sex differences in academic interest and self-concept seem to decline and are related to declines in academic achievement.

This effect seems to be more pronounced for girls than for boys (Fredricks & Eccles, 2002; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Pinxten et al., 2014), although a male advantage in self-concept is present throughout secondary education (Nagy et al., 2010). Overall, the decline in math self-concept and interest might stem from the increasing challenges of math learning compared to other learning subjects (Smith, 2004).

In summary, very little is known regarding whether sex differences in math achievement result from the disparity in the dynamics of self-concept and interest or *vice versa*. This is especially unfortunate since such knowledge would be valuable for the theoretical underpinning of educational interventions that aim to close the gender gap. In addition, the available REM studies neglect the mean level changes between succeeding measurement time points, that is, they are restricted to the analysis of cross-lagged (i.e., the relation between constructs overtime) and stability (i.e., the relation within constructs over time) relations (e.g., see Ganley & Lubienski, 2016; Pinxten et al., 2014).

1.6. The Present Study

In the present study, essentially we examined the sex-related reciprocal effects (or covariance structure) among math achievements (i.e., standardized test-scores and teacher-assigned grades), self-concept, and interest from Grade 5 to 7. More specifically, we used cross-lagged panel analyses across sex groups to study the interplay of both self-concept and interest and achievement. Besides, we extended the examination of sex differences beyond the bivariate relations to the latent mean structure of these constructs and described their trajectories.

2. Method

2.1. Sample

Data were derived from the German longitudinal panel study BiKS (*Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter*; Educational processes, competence development, and selection decisions in preschool and school-age children, see Artelt, Blossfeld, Faust, Rossbach, & Weinert, 2013). The German school system is characterized by an explicit form of school track which is mainly based on prior achievement, more precisely, on teachers' evaluation of students' performance (Maaz, Trautwein, Lüdtke, & Baumert, 2008). In most German federal states (including Bavaria and Hesse), students attend secondary school after 4 years in primary education. The formally different types of secondary school either prepare for apprenticeship (vocational-track and intermediate-track schools) or for university (academic-track-schools).In the present study, we used the second panel of the study (BiKS 8-14), which aims to study the development after the transition to various school types of secondary education (i.e., *Realschule, Hauptschule,* and *Gymnasium*). The total sample of BiKS 8-14 was 3,288 students, of which 1,633 (49.7%) were females. Students were sampled from two German federal states, Bavaria (n = 2,205) and Hesse (n = 1,083).We included all students who took part in the achievement assessment and the motivational questionnaire resulting in an analysis sample of 2,342 students, of which 1,201 were female (51.28%). Students of this sub-sample were followed across three consecutive grades (5, 6, and 7). The following students were excluded from the analysis: Students who repeated a grade (n = 290), attended a comprehensive school type (i.e., *Gesamtschule*, n = 331, as they were not part of the ongoing survey) or a special needs school (*Förderschule, n* = 42), and classes with less than 5 participating students (n = 131), because Marsh & Hau (2003) recommend that the minimal number of students to compute a conclusive school achievement shouldn't be fewer than 4 or 5 students. The average age at Grade 5 was 11.52 years (*SD* = 0.46). The majority of the sample (*n* = 1,481, 63.2%) attended academic track schools (*Gymnasium*), whereas the remaining students (*n* = 861) were enrolled in non-academic track schools (i.e., *Hauptschule* and *Realschule*). Accordingly, students attending academic-track schools were overrepresented. Students came from families with a middle to high socioeconomic status (*M* = 53.17, *SD* = 15.95), based on the *Highest International Socio-Economic Index of Occupation* (HISEI), which has widely been used in the educational large-scale assessment as a proxy for socioeconomic status (Ganzeboom & Treiman, 1996). Both parents were born in Germany for the majority of the sample (*n* = 1,933, 82.5%); 7.8% (*n* = 182) of the students had one parent with a migration background, and for 9.7% (*n* = 227) both parents had a migration background.

2.2. Measures

2.2.1. Math test. Math competence was measured with a newly compiled measure covering the math curricula of Grades 5, 6, and 7 (Klassen, Marx, & Opitz-Karig, 2005; Pekrun et al., 2003). Although the math competence test was designed to measure math in algebra, geometry, and word problems, we used a unidimensional model as suggested by the high correlations between the facets. Most items (88%) had a short answer format; only 12% had a multiple-choice format. In the test development, items were intensively pilot tested and their construct validity confirmed with established measures used in the German PALMA project (*Project for the analysis of competence development in mathematics*; Pekrun et al., 2003). For the three measurement waves (number of items: 44/40/34), an anchor item test design was applied (Frey, Hartig, & Rupp, 2009). Thus, some of the items were unique for the specific grades, while the linking items were identical in adjacent grade-specific test forms. We used plausible values

(PVs) of math competence in the subsequent reciprocal effect model (the procedures of PVs generation presented in the subsequent section). Reliability estimates of math test scores were consistently high in the three waves (.89/.91/.86).

2.2.2. Math grade. Teacher-assigned grades were collected from math teachers' evaluating students' performance at the end of each term. In Germany, grades range from 1 = very good to 6 = unsatisfactory. For ease of understanding, student's grades were recoded, so that higher values indicate better grades.

2.2.3. Math self-concept. Self-concept in math was assessed with the following three items: 1) *Math lessons are easy for me,* 2) *I quickly learn new things in math,* and 3) *I am good in math*. The items are similar to the items previously used to measure the cognitive-evaluative aspect of students' self-concepts (Marsh, 1990; Möller & Pohlmann, 2010). Students were asked to indicate the degree of agreement to these statements on a five-point scale ranging from 1= *not at all*, 2 = *just a little*, 3 = *a fair amount*, 4=*much* and 5 = *very much*. Across grades, the self-concept scale showed high reliability coefficients in terms of McDonald's ω across grades (McDonald, 1999, ω = .92/.94/.94).

2.2.4. Math interest. Subject-specific interest was measured with four items, derived from the German BIJU study (Learning Processes, Educational Careers, and Psychosocial Development in Adolescence and Young Adulthood Study; Baumert et al.,1997), in all grades: 1) *I am looking forward to having a math lesson,* 2) *I would like to have more math lessons than I have right now,* 3) *It is important to me to know a great deal about math,* and 4) *It is important to me to remember information in math*. The first two items assessed the emotional aspect of math interest, while the last two items examined the value aspect (see also Hidi & Renninger, 2006; Krapp, 2002). Similar domain-specific academic interest items, comprising the two aspects of interest (i.e., emotional and value) are commonly used in longitudinal studies (e.g., see Marsh et al., 2005). In the present study, the four items were specified as indicators of the latent construct math interest. The scale showed high reliability estimates (ω = .84/.84/.82).

2.3. Statistical Analysis

Data preparation, recoding, and all analyses were conducted using *R* 3.4.0 (R Development Core Team, 2017); SEM models were estimated with the R package *lavaan* 0.5–23 (Rosseel et al., 2012). The robust maximum likelihood estimator (MLR), which ensure that there are no violations of normality assumptions, was used because the motivational scales had five response categories (Rhemtulla, Brosseau-Liard, & Savalei, 2012). Both math self-concept and interest were modeled on an item level.

2.3.1. Missing data. The percentages of missing values within wave 5 were 3.5% for the math achievement test, 1.4% for the self-concept scale and 1.7% for the interest scale. The missing subjects in subsequent measurement occasions also varied for math competency tests (approx. 12%) and for the motivational variables (approx. 6%). Assuming *missing at random* (MAR, Schafer & Graham, 2002), *Multiple Imputation* was used to handle missing data, providing less biased estimates than traditional approaches such as listwise or pairwise deletion (Enders, 2010). The R package *MICE* (Multivariate Imputation through Chain Equation, Buuren & Groothuis-Oudshoorn, 2011) was used to generate 10 complete data sets. All available background information (i.e., sex, age, school type, socioeconomic status, migration background, math grade, math self-concept, and math interest) were used to set up the imputation model.

Plausible values imputation is a multi-step procedure: First, the data were calibrated separately for Grades 5, 6, and 7, using a unidimensional 2-PL (Two-parameter Logistic) IRT (Item Response Theory) model. Plausible values were drawn for each wave separately. To establish a common scale for all waves, linking was done using the Haberman procedure (Haberman, 2009), which is implemented in the R package *sirt* (Robitzsch, 2017). Second, the plausible values from the first step were used as auxiliary variables in the imputation procedure for missing values on the items of math self-concept and interest, sex, school type, socioeconomic status, migration background, and grades. Third, the plausible values imputation was rerun, using item and discrimination parameter estimates from the first step as fixed values. Additionally, a conditioning or background model (Mislevy, Beaton, Kaplan, & Sheehan, 1992; von Davier, Gonzalez, & Mislevy, 2009) guarantees that it is accounted for all possible interdependencies between math performance competence and conditioning variables (i.e., age, sex, school type, SES, migration background, math grade, self-concept, and interest) in the plausible values. Considering background variables (e.g., sex) within the conditioning model is essential in case group differences (e.g., sex differences) are subsequently analyzed using plausible values (Frey, Hartig, & Rupp, 2009). This complex multi-step procedure is quite common in large-scale assessments (OCED, 2009). In the following analyses, all 10 plausible values were used as manifest indicators of the latent variable score (Frey et al., 2009; Marsh et al., 2005).

2.3.2. Measurement invariance testing across time and sex. In longitudinal studies, measurement invariance (MI) testing across time points is pivotal to make valid statements about the developmental aspect. It is necessary to ensure that the measurement does not change over the course of study (Geiser, 2013). Therefore, we applied longitudinal MI testing (Geiser, 2013; Little, 2013) across time points (Grades 5, 6, and 7) to examine the comparability of measurement instruments for math competence, self-concept, and interest (see further discussion in supplemental materials, appendix A & B). For invariance testing across sex groups, we conducted—a type of the most restrictive form of longitudinal measurement invariance we could establish—a series of increasingly restrictive models within the framework of multiple-group confirmatory factor analysis (Schroeders & Wilhelm, 2011; Wicherts & van Dolan, 2010).

2.3.3. Reciprocal effect model. Structural equation modeling (SEM) was used to address the main research questions. Overall, we specified three separate series of models for each achievement measures (i.e., standardized test scores and teacher-assigned grades). First, we estimated *cross-lagged* and *stability* effects of individual differences in the overall models, which are a full-forwarded model without a group including all three constructs at all three-time points (see Figure 1 & 2, Model 1 & 2, respectively). Second, we modeled *cross-lagged* and *stability* effects of individual differences in math measures, self-concept, and interest from Grade 5 to 7 for the sex group in order to replicate results previously reported for the REM (see Figure 1 & 2, Model 1a & 2a, respectively). Third, to understand the effect of construct overlap and to gauge effects of multicollinearity, we examined the REM in separated models, that is, either math self-concept and achievement (Model 1b & 2b), or math interest and achievement (Model 1c & 2c, details in the supplemental materials, Table S2 & S3). In the sex-specific investigation, the critical test is whether the regression coefficients leading from earlier math self-concept (and interest) to later math achievement and from earlier math achievement to later math self-concept (and interest) across sex groups are identical. That is, if constraining the cross-lagged paths to equality across sex groups leads to significant deterioration in model fit, we conclude that the pattern differs across sexes (Marsh et al., 2005).

To cover the second research question (i.e., sex differences in the mean structure across constructs), we additionally specified a model with correlations among latent constructs; this means, without controlling for prior achievement or motivational constructs. Otherwise, in a regression model, mean differences reflect differences in the residual variances of the constructs. We compared latent group means using multiple-group confirmatory factor analysis (MGCFA), which is superior to more frequently methods relying on manifest scores (Finch & French, 2015). Because of the longitudinal nature of the study and the usage of identical measures for self-concept and interest, we allowed for correlations between item residuals (see also Ganley & Lubienski, 2016; Marsh et al., 2005). Ignoring this correlated uniqueness typically results in positively biased stability coefficients and biased model fit (Marsh & Hau, 1996).

3. Results

3.1. Measurement Invariance Testing Over Time and Sex

Before the main analyses of the study, we ran longitudinal MI and then MGCFA. The longitudinal MI testing procedure consists of a sequence of increasingly restrictive models. The most restrictive form of longitudinal measurement invariance was strict MI, which was supported by the data (i.e., test-scores/grades, self-concept and interest: n = 2,342, χ^2 = 1,406.94, df = 236, CFI = .973; RMSEA = .046, and n = 2,342, χ^2 = 1,407.70, df = 236, CFI = .973; RMSEA = .046, respectively; for a summary of all steps of longitudinal measurement invariance testing see supplemental material Table S1, the upper and lower sections). Subsequently, the model with constraints of strict longitudinal measurement invariance was taken as a baseline model to test for measurement invariance across sex groups. We applied a two-index presentation strategy (Hu & Bentler,1999), that combines an absolute fit index such as *Root Mean Square Error of Approximation* (RMSEA) and an incremental fit index such as *Comparative Fit Index* (CFI), as the *chi-square* statistic is not a decisive indicator of model fit. Also, in case of MI testing across sex groups, the strict measurement invariant model holds (test-scores, self-concept, and interest: n = 2,342, χ^2 = 1,850.76, df = 497, CFI = .969; RMSEA = .048, and grades, self-concept, and interest: n = 2,342, χ^2 = 1,911.95, df = 497, CFI = .968; RMSEA = .049, respectively; for the complete testing procedure see supplemental material Table S1, the upper and lower sections). Accordingly, all subsequent structural equation models were specified with constraints of strict longitudinal MI and strict MI across groups.

3.2. Reciprocal Effects and Stability Among Math Competence, Self-Concept, and Interest

We followed a two-step procedure in the results section as outlined by Marsh et al. (2005). In the first set of analyses, we ran the model with the complete sample to describe the stability and cross-lagged effects between math achievement, self-concept, and interest. In the second set of analyses, we ran multi-group REM analyses (with sex also grouping variable) to examine potential differences in the relations of math achievement, self-concept, and interest. These analyses were done for both achievement measures (i.e., test-scores and teacher-assigned grades).

For the standardized math test-scores, we estimated the overall model (Model 1) and then multi-group REM (Model 1a) including all three constructs at all three-time points (see Figure 1). Both models provided a good fit to the data (see supplemental material Table S1). In Model 1, reciprocal relations between math self-concept and performance could only be found between Grade 6 and 7. Between Grade 5 and 6 only the *self-enhancement effect*, that is, self-concept predicting math test-scores was significant (β = .09**). Taken together, the effects of earlier math self-concept on later math performance was stronger than vice versa for both time points in the overall model (Figure 1). In Model 1a, for female students, early math self-concept positively and significantly predicted later math test-score effects (β = .11* and .18**) over and above the effect of prior math test-scores at both time points. Prior math test-scores, however, positively and significantly predicted later math self-concept (β = .08**) beyond the effect of prior self-concept only at the first time period (see Figure 1). For male students, prior math self-concept positively and significantly predicted subsequent math performance (β = .11*), and vice versa (β = .08*) for the second time point only (i.e., Grade 6 and 7). Considering the standard errors of the parameter estimates these subtle differences between sex groups shouldn't be overemphasized and distract from the overall picture that the coefficients are quite similar.

In both sexes, the comparison of the causal ordering between math test-scores and self-concept more strongly supported the self-enhancement part within the REM model, that is, larger effects of math self-concept on test-scores than vice versa. Model 1 and 1a did not provide support for reciprocal effects between math test-scores and interest; instead, we observed non-significant skill-development and self-enhancement effects (see Figure 1). Stability coefficients were moderately high within the three-year time frame for the overall and group effects (both sexes), that is, for math self-concept (β =.59/.57/.58 to β =.60/.63/.54) and interest (β = .51/.48/.53 and β = .48/.46/.49), whereas the respective coefficients for math test-scores were slightly decreasing (for test-scores: β = .55/.55/.55 to β = .46/.44/.46). A more pronounced stability effect was observed for math self-concept than for interest and test-scores, respectively (see Figure 1).

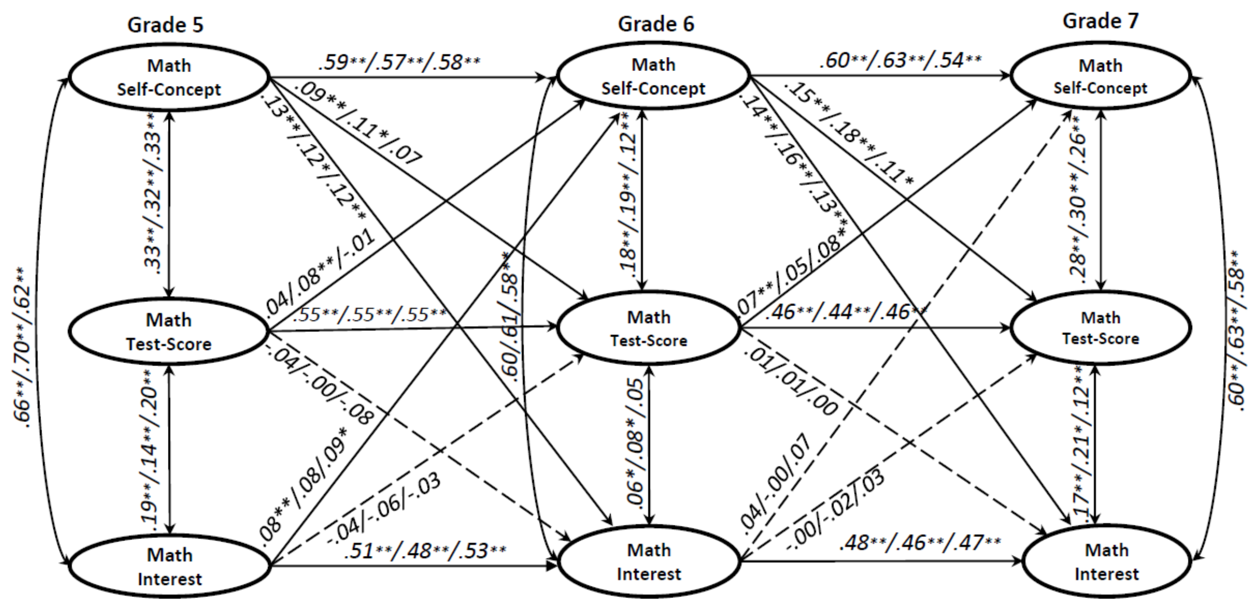


Figure 1. Standardized path coefficients for the REM including achievement, self-concept, and interest using standardized test-scores (Model 1 and Model 1a). *Note*. Overall effects are presented on the left, female effects in the middle, and male effects on the right. Solid paths indicate that results for either overall or female or male effect are statistically significant; dashed paths indicate that results for all are nonsignificant. Correlated uniqueness is considered across time points and sex (not depicted). n = 2,342,**p < .01, * p < .05.

For teacher-assigned grades, we ran the model without grouping (Model 2) and then as a multi-group model (Model 2a) including all three constructs at all three-time points (see Figure 2). Both models provided a satisfactory fit to the data (see supplemental material Table S1). In Model 2, we found the full support of the reciprocal effects between math self-concept and teacher-assigned grades for the first and second time period. Similar to the standardized test-score results, the effects of earlier math self-concept on later math performance were stronger than in the opposite direction at both time points (Figure 2). In Model 2a, for female students, early math self-concept positively and significantly predicted later math test-score effects (β = .13** and β = .18**). Similarly, prior math test-scores positively and significantly predicted later math self-concept (β = .10** and β = .10**) beyond the effect of prior self-concept at all time periods (Figure 2). For male students, prior math self-concept positively and significantly predicted subsequent math performance, and vice versa for the second time period (i.e., Grade 6 and 7), but for the first time period (i.e., Grade 5 and 6) only the self-enhancement effect. Analogous to the standardized test-score interpretation, small differences in the coefficients between the sexes are negligible and within the range of the measurement precision.

Overall, within the REM, the reciprocal relations between teacher-assigned grades and self-concept were more pronounced for the self-enhancement parts than for the skill development effects (Figure 2). The relation between math grade and interest in Model 2 and 2a partially supported negative reciprocal effects in Grade 5 and 6; however, there were no reciprocal relations between for Grade 6 to7. We found stable effects within each domain for the three-year time period and for overall and group effects, that is, for math self-concept (β = .59/.59/.56 to β = .59/.60/.55), interest (β = .51/.51/.51 and β = .50/.51/.45) and grades (β = .58/.58/.58 to β = .49/.47/.51).

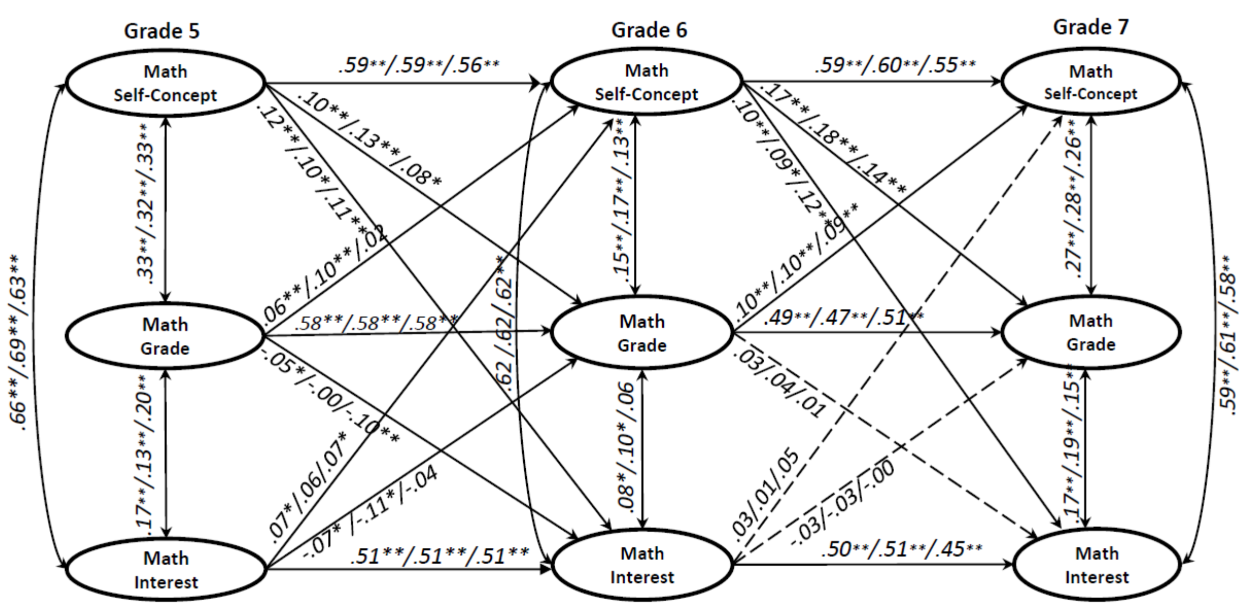


Figure 2. Standardized path coefficients for the REM including achievement, self-concept, and interest using teacher-assigned grades (Model 2 and Model 2a). *Note*. Overall effects are presented on the left, female effects in the middle, and male effects on the right. Solid paths indicate that results for either overall or female or male effect are statistically significant; dashed paths indicate that results for all are nonsignificant. Correlated uniqueness is considered across time points and sex (not depicted). n = 2,342, ** $p < .01$, * $p < .05$.

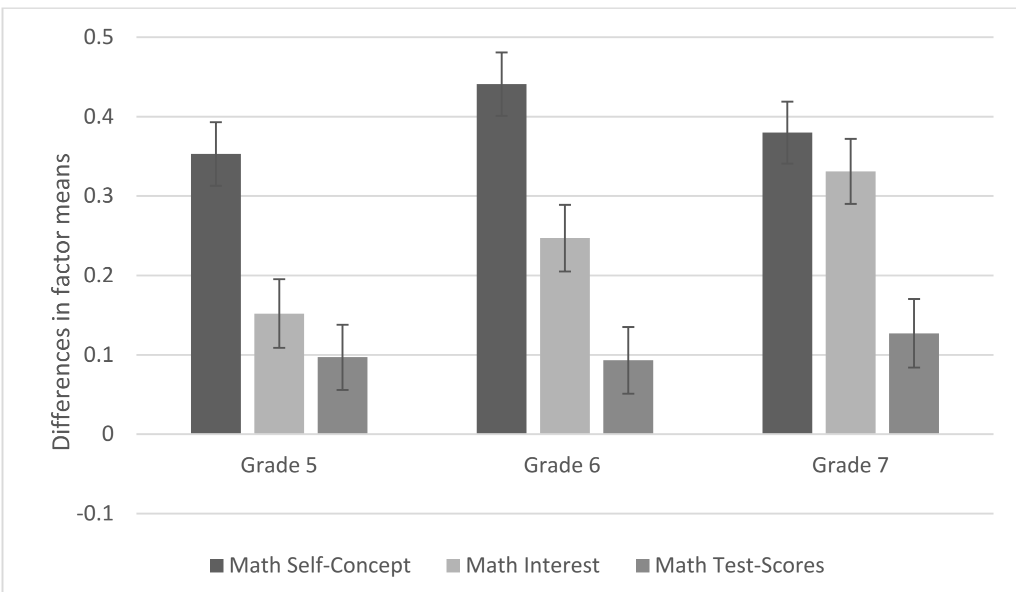
In summary, the *reciprocal relations* between math achievement and self-concept were found for both achievement measures (i.e., test-scores and grades) with a causal predominance of the *self-enhancement effect*. Whereas math self-concept was a consistent predictor of math achievement for both achievement measures, the influence of math achievement on self-concept was inconsistent.

There were substantial positive correlations among the three constructs (e.g., the correlation between math self-concept and interest ranged between $r = .58$ and $r = .70$, in Model 1a & 2a, thus, multicollinearity might obscure the pattern of the results (see also Marsh et al., 2005). For this reason, we conducted a series of models including only one motivational construct at a time. In the "classical" REM with math achievement (i.e., standardized test and grade) and self-concept, even though slightly math teachers' grade vs. self-concept had better *reciprocal relations* than test-scores vs. self-concept (see supplemental material, Table S2 & S3, Model 1b & 2b), overall on both achievement measures the pattern of the results was comparable to the fully specified respective models. That means, the effects were almost similar and the impact of the math interest was negligible. In Model 1c and 2c, with math achievement test and interest, only the effects were small (nearly zero and non-significant, see supplemental material, Table S2 & S3, Model 1c & 2c) for both achievement measures (i.e., test-scores and teacher-assigned grades). Taken together, both Models 1a and 1b, and 3a and 3b did not endorse reciprocal effects between math achievement and interest. Stated differently, prior math self-concept significantly influenced subsequent math interest beyond the effects of earlier measures of math achievement and interest. However, prior math interest had only a small effect on subsequent math self-concept and little or no effect on math competence beyond what could be explained by math self-concept.

3.3. Sex Differences in the Reciprocal Relation and Factor Means

Secondly, we focused on the model with all constructs (Model 1a & 2a) and examined whether the pattern of effects changed across sex groups. That is, we tested for invariance of all regression coefficients across sex. For this purpose, we ran two models: first, the factor loadings, intercepts, and residuals were set to be equal, but the regression coefficients were allowed to vary across sexes (see Model 1a & 2a supplemental material, Table S2 & S3). In a second model (unreported supplementary analyses), in addition to the measurement parameter constraints, the regressions were constrained to be equal across sex groups. To determine whether the changes in the two models were significant or not, we used a chi-square difference test for math standardized test model ($\Delta\chi^2$ (18, N =2,342) = 16.145, p = .56) and for math teachers grade model ($\Delta\chi^2$ (18, N =2,342) = 21.20, p = .27). No significant sex differences were found in relations for both achievement measures. Besides the bivariate relationship given in the REM across sex groups, we also investigated the latent mean structure by comparing the construct means on a latent level in a correlated factor model. Latent means for female students were fixed to zero and freely estimated for the male group; that is, positive values indicating a male advantage. The latent mean differences between males and females in achievement and the motivational constructs over time are given in Figure 3. For the correlational model with standardized test-scores, males possessed a substantially higher math self-concept (*M* = .35/.44/.38) and interest (*M* = .15/.25/.33) across all three-time points, whereas the sex differences for achievement were not substantial (*M* = .10/.09/.13, see also the upper part of Figure 3). For the correlational model with teacher assigned-grades, we found almost identical results: self-concept (*M* = .34/.43/.36), interest (*M* = .12/.23/.31) and math grade (*M* =.10/.08/.12, see the lower part of Figure 3). In other words, males perceived themselves on average as being more competent in math and also viewed themselves as being more interested in math, but this self-evaluation does not correspond to higher math achievement scores.

1: Sex differences in a correlational model with standardized test-scores.



2: Sex differences in a correlational model with teacher-assigned grades.

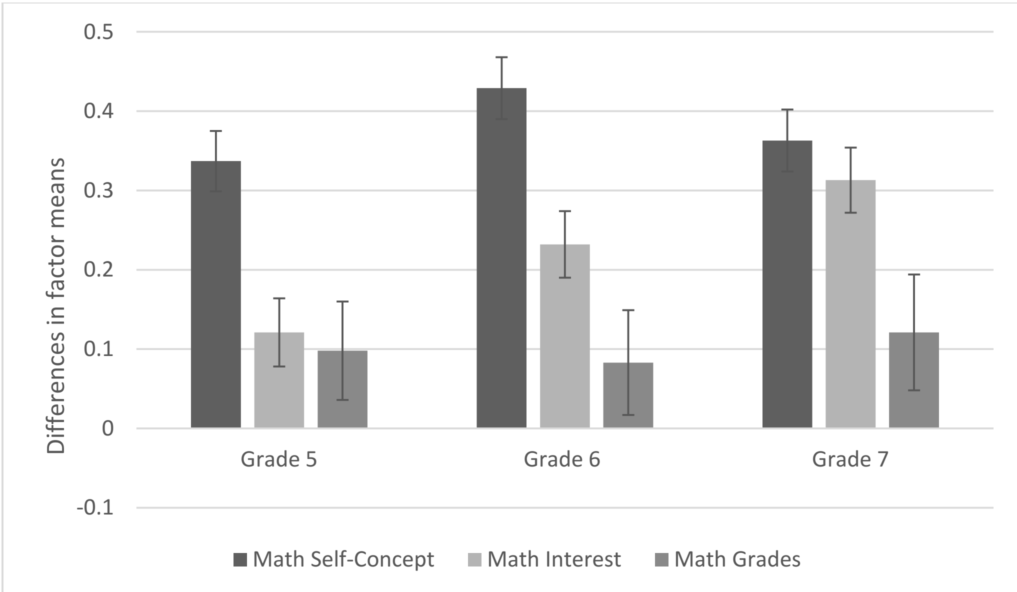


Figure 3. *n* = 2,342. Sex differences in the factor mean based on the correlational models with standardized test and teacher grade. Positive values indicate higher scores for males. Error bars represent standard errors (SE).

4. Discussion

4.1. Reciprocal Relations Among Math Self-Concept, Achievement, and Interest

In the literature on *reciprocal effects*, mixed results have been reported concerning the direction and magnitude of effects between math self-concept and achievement. However, the main models of the present study (Model 1/2, 1a/2a and 1b/2b) evidenced that math self-concept and both achievement measures are *reciprocally* related, that is, *mutually* reinforcing. This result corroborates other empirical findings in the area (e.g., Ganley & Lubienski, 2016; Marsh et al., 1999, 2005; Marsh & Martin, 2011; Niepel et al., 2014; Pinxten et al., 2014). Although the reciprocal effects were positive and significant for both achievement measures, they were more pronounced for grades than for standardized test-scores. This result agrees to the notion that grades are more tangible and salient for students self-beliefs and, therefore, influence self-concepts more strongly than standardized test scores (Marsh et al., 2005; Möller et al., 2014; Pinxten et al., 2010) which do not provide any feedback.

In the present study, we found a non-significant effect of interest on achievement if the self-concept was juxtaposed and not juxtaposed into the model in both achievement measures (Model 1a/2a and 1c/2c). This result challenges the implicit assumption that math interest asserts an effect on achievement in math vice versa. However, previous research using standardized test-scores (e.g., Ganley & Lubienski, 2016; Pinxten et al., 2014) also found non-reciprocal effects between math competence and interest when considered in a model with self-concept simultaneously (see also Marsh et al., 2005). One explanation for the absence of effects is the high construct and empirical overlap between interest and self-concept, which often leads to collinearity issues (see Model 1a /1b vs. Model 1c/2c). Moreover, interest has been shown to indirectly influence school competence (Hübner et al., 2017).

Taken together, the present results provided further evidence for the mutual relationship between math self-concept and achievement (reciprocal effect) for both math achievement measures (i.e., test-scores and grades). This study also evidenced stronger effects of math self-concept on achievement using standardized test-scores and teacher-assigned grades (*self-enhancement effects*). Following the guidelines outlined by Marsh and his colleagues (Marsh & Craven, 2006; Marsh et al., 2005), our study supports the view that math self-concept is a significant and consistent predictor of math achievement, which contrasts with the findings of Ganley and Lubienski (2016) and Pinxten et al. (2014), who reported causal relations (using standardized test-scores) in the opposite direction, that is, math competence is a consistent predictor of self-concept (*skill development effects*).

4.2. Sex Differences in the Reciprocal Relations and Latent Means

We tested whether there are any sex-related differences in the reciprocal relations between variables (i.e., motivational constructs and school achievement). In contrast to the predictions derived from the *differential socialization hypothesis*, which assumes differential paths for the sexes (Marsh, 1989), our results provided strong support for the generalizability of the REM across sex groups. Differently put, sex is not a potential moderator of the reciprocal effects. This finding is in line with previous meta-analytical research (Huang, 2011; Valentine, DuBois, et al., 2004) and recent studies of REM (see Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014) which demonstrated no sex-related difference in the reciprocal relations.

Despite the similarities in the reciprocal relations of motivation and achievement across sex groups, there were substantial sex differences in the mean level of math self-concept and interest in a stereotyped way. Similarly, Marsh et al. (2005) reported substantially higher levels of math self-concept and interest for males. Moreover, in our study, the longitudinal sex differences remain stable for self-concept and seem to increase interest throughout secondary school. This result is in contrast to several studies showing that sex differences seem to decline in academic interest and self-concept (see Arens & Hasselhorn, 2014; Dotterer, McHale, & Crouter, 2009; Fredricks & Eccles, 2002; Jacobs et al., 2002; Pinxten et al., 2014, but see also Nagy et al., 2010, for stable sex differences), which might be due to the specific sample or the cultural context. Two things seem important to stress: First, sex-related mean differences in motivational constructs are crucial because interest and self-concept are more decisive for educational career choices such as the enrollment in STEM subjects in university than potential mean differences in achievement (Hübner et al., 2017; Marsh et al., 2005). Second, the moderate and increasing differences in math self-concept and interest do not transfer to math achievement. Statistically speaking, this is sensible because the mean structure and the variance-covariance structure are independent and, thus, might independently affect students' competence and educational routes.

4.3. Limitations and Implications

Before addressing the implications of our results, some of the shortcomings and strengths of this study should be recognized. First, in our analyses, math competence was modeled as unidimensional competence rather than subdomain-specific (i.e., algebra, geometry, problem-solving), because the dataset did not allow for linking across sub-dimensions. However, as Marsh et al. (1999) noted, having only one or two indicators for a latent construct is rather common in large-scale educational assessments, although this is less than ideal. Second, our findings refer only to the first three years of secondary school; as a result, they cannot be generalized to the entire school time. Third, our sample was drawn from only two German federal states (Bavaria and Hesse). It is known that the bidirectional relations between achievement and motivational variable models vary to some extent across countries and even within the country (Brunner et al., 2013; Marsh & Craven, 2006; Rustemeyer & Fischer, 2005). Thus, the generalizability of our findings to other cultures must be cautiously considered.

Finally, there are different ways to infer "causality" (e.g.. West & Thoemmes, 2010), that is, by associations (the terms cause and effect are often ascribed to variables in a cross-sectional study based on theoretical assumptions), by temporal priority (causes temporally precede effects in a longitudinal study), and by intervention (effects are deliberately manipulated through changing or modifying the assumed cause in a quasi-experimental setting). These different study designs allow for differently strong causal inferences. In the present context, we conceptualized the term “effect” in its orthodox statistical sense and customary path analytical language, that is, indicating a relation that is not necessarily causal (see also Marsh et al., 2017). However, quasi-experimental intervention studies that allow for making strong causal inferences are still sparse (Haney & Durlak, 1998; for an overview see Hattie, 2011). Nonetheless, as a side note, the results of interventions targeting girls’ self-concept in STEM subjects (Dresel & Ziegler, 2006; Ziegler & Heller, 2000) are promising and provide strong evidence for the major role of self-concept and interest in knowledge acquisition and for educational career choices.

With these caveats in mind, our study endorsed the *causal priority* of the academic self-concept overachievement in the context of REM. Please note that the main analyses were conducted after ensuring longitudinal measurement invariance and sex-group-specific measurement invariance. The study replicated REM results previously reported in the literature using standardized test-scores and teacher-assigned grades. That is, we tested the construct with same outcome measures, but different educational context. Such *conceptual replications* of the original findings can be generalized to a larger extent than the *direct replications*, which basically duplicate the sampling, procedures, and assessment (Makel et al., 2012).

The direction of causality among academic self-concept, interest, and achievement has important implications for educators, educational and developmental psychologies who work in school settings (Marsh et al., 2005). Given that our data showed stronger support for the *self-enhancement* hypothesis than the *skill development* hypothesis, teachers and parents should be aware that cultivating and fostering students’ self-concept (interest) is a desirable educational outcome itself, but is also beneficial toward improving students’ achievements. Nevertheless, there is also some evidence for *reciprocal relations* between self-concept and achievement in secondary education, suggesting that improved math self-concepts and interest will lead to better achievements, and improved achievement will result in better math self-concept and interest. Hence, teachers and parents should aim to improve the three constructs simultaneously (see also Hübner et al., 2017).

Even though girls had on overage a lower math self-concept and interest than boys (mean structure), the positive effects of high math self-concept and interest on subsequent math achievement are similar for males and females in the reciprocal relations (variance-covariance structure). Although such patterns might raise concerns about how girls perceive their math abilities and how this systematic underestimation might negatively affect on their girls' future success and potential career paths (Ganley & Lubienski, 2016), having a higher self-evaluation of one's own ability does not necessarily entail better achievement. Therefore, it is necessary to consider both the mean and the variance-covariance structure in future reciprocal relational studies. Because the present study demonstrated no factual differences in a standardized math competence test and teacher-assigned grade, educational efforts to bridge the gender gap should be dedicated toward foster girls' interests in STEM subjects and aim to strengthen their self-concept. However, teachers should do this cautiously as student-teacher interactions in secondary schools tend to favor boys (Drudy & Chatháin, 1998; Robinson & Lubienski, 2011).

4.4. Final Remarks

In conclusion, the present study provided a further endorsement for math self-concept as an important antecedent as well as a significant learning outcome of math achievements (i.e., test-scores and grades). This result strengthens the idea "that self-concept is indeed a 'hot' variable that makes good things happen, facilitating the realization of full human potential in a range of settings " (Marsh & Craven, 2006, p.158). However, in contrast to implicit assumptions, we found no positive effects between math achievement and interest if jointly modeled with self-concept, which is in line with most findings in the research literature (e.g., Marsh et al., 2005). More importantly, psychological constructs are at the heart of our understanding of students' math achievement, and teaching should emphasize the development of stronger motivational constructs. Moreover, the causal interplay between motivational variables and academic achievement is similar for both sexes. Finally, future studies juxtaposing academic achievement, self-concept, and interest in diverse domains (e.g., verbal and science subjects) and different age groups (primary and high school students) need to be conducted to gauge the generalizability of the REM across sexes.

References

Arens, A. K., & Hasselhorn, M. (2014). Age and gender differences in the relation between self-concept facets and self-esteem. *The Journal of Early Adolescence*, *34*, 760–791. doi: 10.1177/0272431613503216.

Artelt, C., Blossfeld, H.-P., Faust, G., Rossbach, H.-G., & Weinert, S. (2013). *Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor-und Grundschulalter* (BiKS-8-14) [Educational processes, competence development and selection decisions in preschool and school age (BiKS-8-14)]. Version: 2. IQB - Institute for Quality Development in Education (Dataset). http://doi.org/10.5159/IQB_BIKS_8_14_v2

Baumert, J, Gruehn, S, Heyn, S, Köller, O, & Schnabel, K. (1997). *Educational process and psycho-social development during adolescence (BIJU)*. Berlin: Max-Planck-Institut für Bildungsforschung.

Bonett, D. G. (2012). Replication-Extension Studies. *Current Directions in Psychological Science*, *21*, 409–412. doi.org/10.1177/0963721412459512

Brunner, M., Gogol, K. M., Sonnleitner, P., Keller, U., Krauss, S., & Preckel, F. (2013). Gender differences in the mean level, variability, and profile shape of student achievement: Results from 41 countries. *Intelligence*, *41*, 378–395. https://doi.org/10.1016/j.intell.2013.05.009

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45,* 1–67. https//doi.org/10.18637/jss.v045.i03

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*, 136–145. https://doi.org/10.1037/0022-0663.69.2.136

Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. *Learning and Individual Differences*, *23*, 172–178. https://doi.org/10.1016/j.lindif.2012.07.021

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9,* 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Dotterer, A. M., McHale, S. M., & Crouter, A. C. (2009). The development and correlates of academic interests from childhood through adolescence. *Journal of Educational Psychology*, *101*, 509–519. https://doi.org/10.1037/a0013987

Dresel, M., & Ziegler, A. (2006). Langfristige Förderung von Fähigkeitsselbstkonzept und impliziter Fähigkeitstheorie durch computerbasiertes attributionales Feedback [Long-term enhancement of academic self-concept and implicit ability theory through computer-based attributional feedback]. *Zeitschrift Für Pädagogische Psychologie*, *20*, 49–63. https://doi.org/10.1024/1010-0652.20.12.49

Drudy, S., & Chatháin, M. U. (1998). Gender differences in classroom interaction in a second level context. *Irish Educational Studies*, *17*, 135–147. https://doi.org/10.1080/0332331980170113

Egorova, M. S., & Chertkova, Y. D. (2016). Sex differences in mathematical achievement: Grades, national test, and self-confidence. *Psychology in Russia*, *9*, 4–23. https://doi.org/10.11621/pir.2016.0301

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, *136*, 103–127. https://doi.org/10.1037/a0018053

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

Finch, W. H., & French, B. F. (2015). *Latent Variable Modeling with R*. New York: Routledge.

Fraine, B. D., Damme, J. V., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, *32*, 132–150. https://doi.org/10.1016/j.cedpsych.2006.10.005

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. https://doi.org/10.1037/0012-1649.38.4.519

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics —A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*, 497–514. https://doi.org/10.1007/BF03173468

Frey, A., Hartig, J., & Rupp, A. A. (2009). An NCME Instructional Module on Booklet Designs in Large-Scale Assessments of Student Achievement: Theory and Practice. *Educational Measurement: Issues and Practice*, *28*, 39–53. https://doi.org/10.1111/j.1745-3992.2009.00154.x

Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, *47*, 182–193. https://doi.org/10.1016/j.lindif.2016.01.002

Ganzeboom, H. B., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, *25*, 201–239.

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J.-P., … Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development*, *87*, 165–175. https://doi.org/10.1111/cdev.12458

Geiser, C. (2013). *Data Analysis with Mplus*. New York: Guilford Press.

Gibbs, B. G. (2010). Reversing fortunes or content change? Gender gaps in math-related skill throughout childhood. *Social Science Research*, *39*, 540–569. https://doi.org/10.1016/j.ssresearch.2010.02.005

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

Haberman, S. J. (2009). Linking parameter estimates derived from an item response model through separate calibrations. *ETS Research Report Series*, *2009*, i-9. https://doi.org/10.1002/j.2333-8504.2009.tb02197.x

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*, 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Halpern, D., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, *12*, 229-246. https://doi. org/10.1023/A:1009027516424.

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

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*, 111–127. https://doi.org/10.1207/s15326985ep4102_4

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. https://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. https://doi.org/10.1016/j.jsp.2011.07.001

Hübner, N., Wille, E., Cambria, J., & Trautwein, U. (2017). Maximizing gender equality in STEM by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM. *Journal of Educational Psychology,* 109, 993-1009. https://doi.org/10.1037 / edu0000183.

Hyde, J. S. (2005). The gender similarities hypothesis. *The American Psychologist*, *60*, 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Hyde, J., S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, *107*, 139–155.

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. https://doi.org/10.1111/1467-8624.00421

Jansen, M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology*, *46*, 116–127. https://doi.org/10.1016/j.cedpsych.2016.05.004

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, *32*, 448–470. https://doi.org/10.2307/749801

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*, 383–409. https://doi. org/10.1016/S0959-4752(01)00011-1

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, *31*, 357–365. https://doi.org/10.1177/0165025407077757

Little, T. (2013). *Longitudinal Structural Equation Modeling*. Guilford Press.

Liu, O. L. (2009). An investigation of factors affecting gender differences in standardized math performance: Results from U.S. and Hong Kong 15 year olds. *International Journal of Testing*, *9*, 215–237. https://doi.org/10.1080/15305050903106875

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, *2*, 99–106. https://doi.org/10.1111/j.1750-8606.2008.00048.x

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*, 537–542. https://doi.org/10.1177/1745691612460688

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*, 417–430. https://doi.org/10.1037/0022-0663.81.3.417

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

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, *34*, 155–167. https://doi.org/10.1207/s15326985ep3403_2

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. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit. *The Journal of Experimental Education*, *64*, 364–390. https://doi.org/10.1080/00220973.1996.10806604

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. *The American Psychologist*, *58*, 364–376. https://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., Hau, K.-T., & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with

Chinese) for Hong Kong students. *American Educational Research Journal*, *39*, 727–763. https://doi/10.3102/00028312039003727.

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

Marsh, H. W., Pekrun, R., Murayama, K., Arens, K. A., Parker, P. D., Guo, J., & Dicke, T. (2017). An integrated model of academic self-concept development: academic self-concept, grades, test scores, and tracking over six years. *Developmental Psychology, 54*, 263–280. https://doi.org/10.1037/dev0000393

Marsh, H. W., 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. https://doi.org/10.1111/j.1467-8624.2005.00853.x

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

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, *35*, 705–738. https://doi.org/10.3102/00028312035004705

Marx, H., & Opitz-Karig, U. (2005). *Deutscher Mathematiktest für 5. Klassen (DEMAT 5+)*. [German Mathematics Test for 5th grade (DEMAT 5+)]. Unpublished: University of Leipzig.

McDonald, R. P. (1999). *Test theory: A unified treatment*, Mahwah, N. J.: Erlbaum Associates.

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

Möller, J., & Pohlmann, B. (2010). Achievement differences and self-concept differences: Stronger associations for above or below average students? *British Journal of Educational Psychology*, *80*, 435–450. https://doi.org/10.1348/000709909X485234

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48, 1315-1346. https://doi.org/10.3102/0002831211419649

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, *84*, 591–611. https://doi.org/10.1111/bjep.12047

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. https://doi.org/10.3102/0034654309337522

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students’ mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence*, *20*, 482–506. https://doi.org/10.1111/j.1532-7795.2010.00644.x

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*, 1170–1191. https://doi.org/10.1037/a0036307

OCED. (2009). *PISA 2006 Technical report*. Paris: OECD.

Pekrun, R., Götz, T., Jullien, S., Zirngibl, A., von Hofe, R., & Blum, W. (2003). *Skalenhandbuch PALMA 2.* Messzeitpunkt (6. Klassenstufe) [Documentation of the PALMA 2]. University of Munich: Institute of Educational Psychology.

Pinxten, M., de Fraine, B., van Damme, J., & D'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. *British Journal of Educational Psychology*, *80*, 689–709. https://doi.org/10.1348/000709910X493071

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, *84*, 152–174. https://doi.org/10.1111/bjep.12028

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354–373. https://doi.org/10.1037/a0029315

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*, 268–302. https://doi.org/10.3102/0002831210372249

Robitzsch, A. (2017). *sirt: Supplementary item response theory models* (R package version 2.025). https://CRAN.R-project.org/package=sirt

Rosseel, Y. (2012). *lavaan:* An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*, 1-36. http://www.jstatsoft.org/v48/i02/

Rustemeyer, R., & Fischer, N. (2005). Sex- and age-related differences in mathematics. *Psychological Reports*, *97*, 183–194. https://doi.org/10.2466/pr0.97.1.183-194

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67. https://doi.org/10.1006/ceps.1999.1020

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

Schroeders, U., & Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, *71*, 849–869. https://doi.org/10.1177/0013164410391468

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, *67*, 135–149. https://doi.org/10.1080/00220979909598349

Smith, A. (2004). Making mathematics count: the report of Professor Adrian Smith’s inquiry into post-14 mathematics education. http://dera.ioe.ac.uk/id/eprint/4873.

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

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

Valentine, J. C., & Dubois, D. L. (2005). Effects of self-beliefs on academic achievement and vice versa. Separating the chicken from the egg. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self-research: New frontiers for self-research* (Vol. 2, pp. 53-78). Greenwich, CT: Information Age.

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

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. https://doi.org/10.1177/109442810031002

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

von Maurice, J., Dörfler, T., & Artelt, C. (2014). The relation between interests and grades: Path analyses in primary school age. *International Journal of Educational Research*, *64,* 1–11. https://doi.org/10.1016/j.ijer.2013.09.011

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th through 11th-Grade Australian students. *Child Development*, *75*, 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*, 18–37. https://doi.org/10.1037/a0015917

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, *29*, 39–47. https://doi.org/10.1111/j.1745-3992.2010.00182.x

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*, 49–78. https://doi.org/10.1007/BF02209024

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

Yoon, K. S, Eccles, J. S, & Wigfield, A. (1996). Self-concept of ability, value, and academic achievement: A test of causal relations. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1996. Washington: American Educational Research Association.

Ziegler, A., & Heller, K. A. (2000). Effects of an attribution retraining with female students gifted in physics. *Journal for the Education of the Gifted*, *23*, 217–243. https://doi.org/10.3102/00028312039003727

Supplemental materials

Appendix A: Measurement invariance testing across time and sex

Appendix B: Results of measurement invariance testing over time and sex

Table S1: The goodness of fit indices for longitudinal and multi-group MI testing

Table S2: Standardized path coefficients using standardized test-scores and motivational variables for main model 1a, 1b, and 1c

Table S3: Standardized path coefficients using teacher-assigned grades and motivational variables for main model 2a, 2b, and 2c

Appendix A: Measurement Invariance Testing Across Time and Sex

In longitudinal studies, measurement invariance (MI) testing across time points is pivotal to make valid statements about the developmental aspect. It is necessary to ensure that the measurement does not change over the course of study (Geiser, 2013). Therefore, we applied longitudinal MI testing (Geiser, 2013; Little, 2013) across time points (Grades 5, 6, and 7) to examine the comparability of measurement instruments for math competence, self-concept, and interest. In the baseline model, all measurement parameters (i.e., factor loadings, intercepts, and residual variances) were freely estimated across measurement occasions (*configural model*), only the residuals of similarity worded items were allowed to correlate across time points. Factor means were fixed at the first time point and freely estimated at all other time points. For *metric invariance,* the factor loadings were fixed to be equal across time points. In addition, intercepts are fixed to be equal in *scalar invariance* testing. In the most restrictive form (*strict invariance*), all measurement parameters are equal across time points. In comparison to MI testing across sex groups, which is explained in the next paragraph, in longitudinal MI, the residual of similarity worded items is allowed to co-vary across measurement occasions.

For invariance testing across sex groups, we conducted—a type of the most restrictive form of longitudinal measurement invariance we could establish—a series of increasingly restrictive models within the framework of multiple-group confirmatory factor analysis (Schroeders & Wilhelm, 2011; Wicherts & van Dolan, 2010). In the least restrictive model of *configural measurement invariance*, all measurement parameters (i.e., factor loadings, intercepts, and residuals) vary freely in each group. In subsequent steps of measurement invariance testing, additional equality restrictions are imposed on specific parameters across sex groups as follows: *metric invariance* (factor loadings), *strong invariance* (factor loadings and intercepts), and *strict invariance* (factor loadings, intercepts, and residuals). To make this point clear, in longitudinal MI testing, restrictions are introduced across measurement time points; in MI testing across sex groups restrictions are introduced across sex groups. Because the complete MI testing comprises of 16 different models in total, we tested in a two-step procedure. To assess whether imposing additional constraints on measurement parameters leads to deterioration in model fit, the differences in the CFI between two consecutive models was evaluated (Vandenberg & Lance, 2000), whereas a deterioration in a model of CFI $\geq .01$ is usually considered indicative of a serious deterioration.

Appendix B: Results of Measurement Invariance Testing Over Time and Sex

We executed two measurement invariances on the basis of math test score and teacher grade. First, for the math test score and motivational variables, the longitudinal MI testing consists of a sequence of increasingly restrictive models (see Table 1, upper part). The most restrictive form of longitudinal measurement invariance was strict MI. Subsequently, this strict form of measurement invariance was taken as a baseline model to test for measurement invariance across sex groups. In more detail, in the step of *configural invariance*, all factor loadings, intercepts, and residuals were freely estimated, only the overall structure was supposed to be identical across groups. It should be noted that although the measurement parameters are allowed to freely vary *across groups*, the restrictions of longitudinal measurement invariance hold; that is, factor loadings, intercepts, and residuals are fixed to equality *across time points* (see Table 1, lower part). According to the cut-off values reported by Hu & Bentler (1999), model fit indices of the configural model were good: n = 2,342, χ^2 = 1,718.06, df = 485, p < .000; CFI = .972; RMSEA = .047. In the second step, *metric invariance*, only factor loadings were constrained to be equal across sex groups; means were fixed to zero. Additionally imposing the equality constraints of metric invariance did not yield a meaningful deterioration in the model fit (n = 2,342, χ^2 = 1,749.34, df = 490, p < .001; CFI =. 971; RMSEA = .047, ΔCFI = .001, ΔRMSEA = .000). In the third step, *strong invariance* (also labeled *scalar invariance*), factor loadings and intercepts were constrained to be equal across the sex groups, and residuals were freely estimated, which did not result in significant changes in the model fit: n = 2,342, χ^2 = 1,785.07, df = 491, p < .001; ΔCFI = .001, ΔRMSEA = .000). In addition to the constraints of scalar invariance, in the last step of measurement invariance testing (*strict invariance*) the residuals were also fixed to equality across groups. Even this model held: n = 2,342, χ^2 =

1,850.76, *df* = 497, $p < .001$; ΔCFI = .001, ΔRMSEA = .001). Accordingly, all subsequent structural equation models were specified with constraints of strict longitudinal MI and strict MI across groups. Second, for the math teacher grade and motivational variables, we exactly repeated the same procedure and the detailed results were present in the lower part of Table S1.

Table S1. The goodness of fit indices for longitudinal and multi-group MI testing

longitudinal MI testing with standardized test score and motivational variables							
Models (complete sample)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	1,179.79	204	. 978	–	.970	.045	
Metric invariance	1,226.93	214	. 977	.000	.970	.045	.000
Strong invariance	1,306.18	222	. 975	.002	.969	.046	.001
Strict invariance	1,406.94	236	. 973	.002	.969	.046	.000
Testing MI across sex groups for the strict longitudinal MI model							
Models (MGCFA)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	1,718.06	485	. 972		.968	.047	
Metric invariance	1,749.34	490	. 971	.001	.967	.047	.000
Strong invariance	1,785.06	491	. 970	.001	.966	.047	.000
Strict invariance	1,850.76	497	. 969	.001	.965	.048	.001
longitudinal MI testing with teacher-assigned grade and motivational variables							
Models (complete sample)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	1,193.98	204	. 978	–	.970	.046	
Metric invariance	1,242.19	214	. 977	.001	.970	.045	.001
Strong invariance	1,316.13	222	. 975	.002	.969	.046	.001
Strict invariance	1,407.70	236	. 973	.002	.969	.046	.000
Testing MI across sex groups for the strict longitudinal MI model							
Models (MGCFA)	χ^2	*df*	CFI	ΔCFI	TLI	RMSEA	ΔRMSEA
Configural invariance	1,791.59	485	. 970		.966	.048	
Metric invariance	1,823.95	490	. 969	.001	.966	.048	.000
Strong invariance	1,862.67	491	. 969	.000	.965	.049	.001
Strict invariance	1,911.95	497	. 968	.001	.964	.049	.000

Note. n = 2,342; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Table S2. Standardized path coefficients for different variants of the REM (standardized test-scores)

Variables	Model 1a		Model 1b		Model 1c	
Cross-lagged paths	Females	Males	Females	Males	Females	Males
Grade 5 → Grade 6						
Msc → Mcom	.11 (.02)*	.07 (.02)	.06 (.03)	.05 (.03)		
Mcom → Msc	.08 (.03)**	-.01(.03)	.08 (.03)*	-.00 (.03)		
Mcom → Mint	-.01 (.03)	-.06 (.03)			.02 (.03)	-.05 (.03)
Mint → Mcom	-.00 (.04)	-.08 (.03)			.00 (.03)	.01 (.03)
Msc → Mint	.12 (.05)*	.12 (.04)**				
Mint → Msc	.08 (.04)	.09 (.04) *				
Grade 6 → Grade 7						
Msc → Mcom	.18 (.02)**	.11(.02)*	.16 (.03)***	.12 (.03)***		
Mcom → Msc	.05 (.03)	.08 (.03)*	.05 (.03)	.07 (.03)*		
Mcom → Mint	.01 (.03)	.00 (.03)			.04 (.03)	.03 (.03)
Mint → Mcom	-.02 (.04)	.03 (.04)			.10 (.03)**	.09 (.03)**
Msc → Mint	.16 (.04)***	.13 (.04)**				
Mint → Msc	-.00 (.04)	.07(.04)				
Stability paths						
Grade 5 → Grade 6						
Msc → Msc	.57 (.04) ***	.58 (.03) ***	.63 (.02) ***	.63 (.02) ***		
Mcom → Mcom	.55 (.02) ***	.55 (.02) ***	.56 (.02) ***	.55 (.02) ***	.58 (.02) ***	.57 (.02) ***
Mint → Mint	.48 (.03) ***	.53 (.04) ***			.59 (.02) ***	.60 (.02) ***
Grade 6 → Grade 7						
Msc → Msc	.63 (.05) ***	.54 (.04) ***	.63 (.02) ***	.59 (.02) ***		
Mcom → Mcom	.44 (.02) ***	.46 (.02) ***	.44 (.02) ***	.46 (.02) ***	.48 (.02) ***	.55 (.02) ***
Mint → Mint	.46 (.04) ***	.49 (.04) ***			.57 (.02) ***	.54 (.02) ***

Fit indices	Model 1a	Model 1b	Model 1c
χ^2	1,850.76	326.59	1,192.55
Df	497	111	186
CFI	.969	.990	.953
TLI	.965	.988	.946
RMSEA	.048	.041	.068

Note. Mcom= math competence; Msc= math self-concept; Mint= math interest. Regression coefficients and standard errors β (*SE*) for each sex. $n = 2{,}342$, CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S3. Standardized path coefficients for different variants of the REM (teacher-assigned grades)

Variables	Model 2a		Model 2b		Model 2c	
Cross-lagged paths	Females	Males	Females	Males	Females	Males
Grade 5 → Grade 6						
Msc → Mcom	.13 (.04)**	.08 (.03)*	.05 (.03)	.06 (.03)		
Mcom → Msc	.10 (.03)***	.02 (.03)	.10 (.03)***	.02 (.03)		
Mcom → Mint	-.00 (.03)	-.10 (.03)**			.03 (.03)	-.07 (.03)*
Mint → Mcom	-.11 (.04)*	-.04 (.03)			-.02 (.03)	.01 (.03)
Msc → Mint	.10 (.05)*	.11 (.04)**				
Mint → Msc	.06 (.04)	.07 (.04) *				
Grade 6 → Grade 7						
Msc → Mcom	.18 (.04)***	.14(.04)**	.17 (.03)***	.14 (.03)***		
Mcom → Msc	.10 (.03)***	.09 (.03)**	.10 (.03)***	.09 (.03)**		
Mcom → Mint	.04 (.03)	.01 (.03)			.06 (.03)	.04 (.03)
Mint → Mcom	-.03 (.04)	-.00 (.04)			.10 (.03)**	.09 (.03)*
Msc → Mint	.09 (.04)*	.12 (.04)**				
Mint → Msc	.01 (.04)	.05 (.04)				
Stability paths						
Grade 5 → Grade 6						
Msc → Msc	.59 (.04) ***	.56 (.03) ***	.62 (.02) ***	.60 (.02) ***		
Mcom → Mcom	.58 (.02) ***	.58 (.02) ***	.59 (.02) ***	.58 (.02) ***	.61 (.02) ***	.60 (.02) ***
Mint → Mint	.51 (.05) ***	.51 (.03) ***			.58 (.02) ***	.57 (.02) ***
Grade 6 → Grade 7						
Msc → Msc	.60 (.05) ***	.55 (.04) ***	.61 (.02) ***	.58 (.02) ***		
Mcom → Mcom	.47 (.02) ***	.51 (.02) ***	.48 (.02) ***	.51 (.02) ***	.51 (.02) ***	.54 (.02) ***
Mint → Mint	.51 (.04) ***	.45 (.04) ***			.58 (.02) ***	.3 (.02) ***

Fit indices	Model 2a	Model 2b	Model 2c
χ^2	1,911.95	375.31	1,158.71
df	497	111	186
CFI	.968	.968	.955
TLI	.964	.964	.949
RMSEA	.049	.049	.067

Note. Mcom= math competence; Msc= math self-concept; Mint= math interest. Regression coefficients and standard errors β (*SE*) for each sex. $n = 2,342$, CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation. *** $p < .001$, ** $p < .01$, * $p < .05$