The Reciprocal 2I/E Model: An Investigation of Mutual Relations Between Achievement and Self-Concept Levels and Changes in the Math and Verbal Domain Across Three Countries

Fabian Wolff¹, Fabio Sticca², Christoph Niepel³, Thomas Götz⁴, Jan Van Damme⁵,
and Jens Möller¹

¹Kiel University   ²Marie Meierhofer Institute for the Child   ³University of Luxembourg

⁴University of Vienna   ⁵KU Leuven

First submitted on 28th of January, 2020

Re-submitted with major revisions on 27th of April, 2020

Re-submitted with minor revisions on 5th of July, 2020

Author Note

Fabian Wolff (fwolff@ipl.uni-kiel.de; +49 431 880 4964) and Jens Möller (jmoeller@ipl.uni-kiel.de; +49 431 880 1241) are at the Institute for Psychology of Learning and Instruction, Kiel University, Olshausenstrasse 75, 24118 Kiel, Germany. Fabio Sticca (sticca@mmi.ch; +44 205 5220) is at the Marie Meierhofer Institute for the Child, Pfingsweidstraße 16, 8005 Zürich, Switzerland. Christoph Niepel (christoph.niepel@uni.lu; +352 4666 446747) is at the Institute of Cognitive Science and Assessment, University of Luxembourg, 11 Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg. Thomas Götz (thomas.goetz@univie.ac.at; +43 14277 47382) is in the Faculty of Psychology, University of Vienna, Universitätsstraße 7, 1010 Wien, Austria. Jan Van Damme (jan.vandamme@kuleuven.be; +32 163 26245) is in the Faculty of Psychology and Educational Sciences, KU Leuven, Dekenstraat 2, 3000 Leuven, Belgium.

This research was funded by grants from the German Research Foundation (DFG) to Fabian Wolff (WO 2341/1-1) and Jens Möller (MO 648/15-1/15-3), from the Luxembourg National Research Fund (FNR) to Christoph Niepel (C16/SC/11333571), from the Swiss National Science Foundation (SNSF) to Thomas Götz (100014_131713/1), and from the Flemish Government to the Center of Educational Effectiveness and Evaluation, KU Leuven (through the Policy Research Center “Educational and School Careers”).

We are indebted to Nathalie Vandenberghe for her help with data preparation in Study 3. We thank Stephen McLaren and Vivien Bogdanski for copyediting the manuscript.

The data used in this research can be found here: <https://osf.io/7f6gb/>

Please address correspondence concerning this article to Fabian Wolff.

The Reciprocal 2I/E Model: An Investigation of Mutual Relations Between Achievement and Self-Concept Levels and Changes in the Math and Verbal Domain Across Three Countries

First submitted on 28th of January, 2020

Re-submitted with major revisions on 27th of April, 2020

Re-submitted with minor revisions on 5th of July, 2020

Abstract

Two longitudinal extensions of the classic I/E model have attracted researchers' attention in recent years: The reciprocal I/E model (RI/EM) describes the reciprocal effects between students' math and verbal achievements and self-concepts. The 2I/E model (2I/EM) describes the effects of students' math and verbal achievement levels and changes on their math and verbal self-concepts. The present research integrates these two approaches into the reciprocal 2I/E model (R2I/EM), which describes the effects of students' math and verbal achievement levels and changes on their math and verbal self-concepts and the effects of their math and verbal self-concept levels and changes on their math and verbal achievements. We tested this model in three empirical studies, including about 6,000 students between Grade 4 and Grade 11 from Germany, Switzerland, and Belgium, using grades and test scores as achievement indicators. Overall, we found strong support for the R2I/EM. In line with prior 2I/EM studies, students' achievement levels showed positive (negative) effects on their self-concepts within (between) subjects. The effects of students' achievement changes on their self-concepts within (between) subjects were usually positive (near-zero). In line with prior RI/EM studies, students' self-concept levels showed positive (negative or near-zero) effects on their achievements within (between) subjects. For the first time, we also examined the effects of students' self-concept changes on their achievements: These effects were usually positive (positive or near-zero) within (between) subjects. Our findings have important theoretical and practical implications for the I/E model literature and for teachers aiming to foster their students' self-concepts and achievements.

Keywords: academic self-concept, dimensional comparisons, I/E model, social comparisons, temporal comparisons

Educational Impact and Implications Statement

Previous research has shown that students' subject-specific self-concepts are formed by the joint operation of three comparison processes: social comparisons (where students compare their achievements with their classmates' achievements), dimensional comparisons (where students compare their own achievements in different subjects), and temporal comparisons (where students compare their own achievements across time). The findings of the present research suggest that students conduct comparisons not only of their subject-specific achievements, but also of their subject-specific self-concepts, with resulting impacts on their subsequent achievements: In particular, we found positive effects of changes in students' self-concept in one subject (i.e., temporal comparisons of self-concepts) on their subsequent achievements in the same and other subjects. This finding is of particular practical importance, as it demonstrates that teachers can help students enhance their achievements in different subjects via impacting on their self-perceptions (i.e., fostering students' self-concepts and supporting students to become aware of increases in their self-concepts).

The Reciprocal 2I/E Model: An Investigation of Mutual Relations Between Achievement and Self-Concept Levels and Changes in the Math and Verbal Domain Across Three Countries

Previous research has shown that students' academic self-concepts—understood as mental representations of their academic abilities in different subjects—are formed not only on the basis of students' prior achievements but also affect their subsequent achievements to a substantial degree (e.g., Marsh, Seaton, Dicke, Parker, & Horwood, 2019). The reciprocal internal/external frame of reference model (RI/EM; Möller, Retelsdorf, Köller, & Marsh, 2011) describes these mutual relations between subject-specific achievements and self-concepts. It combines the reciprocal effects model (REM; Marsh & Craven, 2006), which integrates the effects of achievements on self-concepts and of self-concepts on achievements in single subjects, and the classic I/E model (I/EM; Marsh, 1986), which only considers the effects of achievements on self-concepts, but includes both math and verbal subjects. Thus, the RI/EM also shares the core assumption of the I/EM that students form their subject-specific self-concepts as a result of the joint operation of two comparison processes: social comparisons (where students compare their achievement in one subject with their classmates' achievements in the same subject; Festinger, 1954) and dimensional comparisons (where students compare their math and verbal achievements; Möller & Marsh, 2013).

However, recent research on the impact of comparison processes (e.g., Müller-Kalthoff, Helm, & Möller, 2017; Wolff, Helm, Zimmermann, Nagy, & Möller, 2018) has shown that temporal comparisons (where students compare their current achievement with their prior achievement in one subject; Albert, 1977) also affect students' self-concepts, along with social and dimensional comparisons. Therefore, Wolff and colleagues (Wolff, Helm, & Möller, 2019; Wolff, Nagy, et al., 2019) developed the 2I/E model (2I/EM), as another longitudinal extension of the I/EM, in which all three comparison processes are considered; one external (social) and two internal (dimensional and temporal). Similarly to the I/EM, the 2I/EM considers the effects of achievements on self-concepts only, while neglecting the

reverse effects of self-concepts on achievements. Thus, it is still unclear whether students' achievements are affected not only by their self-concept levels (as shown in studies testing the RI/EM), but also by changes in their self-concepts (i.e., temporal comparisons of self-concepts). Such knowledge would be of great theoretical interest for understanding the psychological processes associated with students' self-concept development, as well as with regard to the generalizability of comparison processes in the educational context.

Furthermore, it could have important practical implications by pointing out new strategies for teachers to help students improve their achievements via impacting on their self-perceptions.

The present paper addresses this research gap. To this end, we develop the reciprocal 2I/E model (R2I/EM) by integrating the RI/EM and the 2I/EM. Subsequently, we test the R2I/EM in three empirical studies, using student samples from three different countries (Germany, Switzerland, Belgium), and compare the effects of students' achievement levels and achievement changes on their self-concepts with the effects of their self-concept levels and self-concept changes on their achievements.

Theoretical Background

The Reciprocal Effects Model

There has been a long debate in educational research about the causal ordering of academic achievement and self-concept (e.g., Marsh, 1990a; Marsh & Craven, 2006). Calsyn and Kenny (1977) distinguished between the skill-development model and the self-enhancement model. The skill-development model suggests that achievement has a positive influence on subsequent self-concept. For example, this can be explained by social comparisons (e.g., Möller & Pohlmann, 2010): As high (low) performing students have plenty of opportunities to compare their achievements with those of their lower (higher) performing classmates and thus frequently experience their achievement as being above-average (below-average), they will develop higher (lower) self-concepts (e.g., "I have the best math grade in my class, so I must have very high math abilities"). In contrast, the self-enhancement model

assumes that self-concept impacts positively on subsequent achievement. Following social cognition theory (Bandura, 1986), this may be due to the fact that a higher (lower) self-concept encourages (discourages) students to engage in subject-relevant activities. In consequence of higher (lower) engagement, effort, and persistence, students will enhance (worsen) their achievement in this subject (e.g., Marsh & Yeung, 1997b).

Marsh and Craven (2006) integrated the skill-development and self-enhancement models into their REM by positing that achievement and self-concept are mutually reinforcing. As illustrated in Figure 1, the REM makes three central assumptions: First, there should be positive effects of achievement on subsequent self-concept, even after controlling for prior self-concept (skill-development effects; α). Second, there should be positive effects of self-concept on subsequent achievement, even after controlling for prior achievement (self-enhancement effects; β). Third, there should be positive effects of achievement on subsequent achievement as well as of self-concept on subsequent self-concept that reflect the stability of these constructs across time (autoregressive effects; ϵ). Empirical studies have provided strong support for these assumptions in different subjects and across different age groups. However, the skill-development effects often tended to be somewhat stronger than the self-enhancement effects (e.g., Guay, Marsh, & Boivin, 2003; Marsh & Craven, 2006; Marsh & Martin, 2011; Marsh & O'Mara, 2008; Marsh & Yeung, 1997a; Retelsdorf, Köller, & Möller, 2014; Valentine, DuBois, & Cooper, 2004) and the results were not always consistent if students in their first school years were examined (see Ehm, Hasselhorn, & Schmiedek, 2019; Weidinger, Steinmayr, & Spinath, 2018, for an overview).

The Classic I/E Model

Unlike the REM, the I/EM only considers the effects of student' achievements on their self-concepts, but takes into consideration both math and verbal subjects (see Figure 2). Originally, Marsh (1986) developed this model to explain the seemingly paradoxical finding that the correlation between students' math and verbal self-concepts is usually close to zero,

although their math and verbal achievements are generally strongly positively correlated. He argued that this difference in achievement and self-concept correlations might be due to the simultaneous operation of social and dimensional comparisons: On the one hand, students compare their achievement in one subject with their classmates' achievements in the same subject, leading to strong positive effects of achievements on self-concepts within subjects in the I/EM (social comparison effects; α)—similar to the skill-development effects in the REM. On the other hand, students compare their math and verbal achievements with each other, leading to moderate negative effects of achievements on self-concepts between subjects in the I/EM (dimensional comparison effects; γ). More recent research suggests that these dimensional comparison processes are triggered in particular by students' motivation to find out their strengths and weaknesses as part of their search for identity (Möller, Helm, Müller-Kalthoff, Nagy, & Marsh, 2015; Wolff, Helm, & Möller, 2018).

Empirically, more than 100 studies have provided support for the I/EM (e.g., Marsh & Hau, 2004; Möller, Wolff, & Helm, in press; Wolff, Helm, Junge, & Möller, 2020; Wolff, Nagy, Helm, & Möller, 2018). For example, Möller, Pohlmann, Köller, and Marsh (2009) found strong social comparison effects and moderate negative dimensional comparison effects in a meta-analysis integrating 69 data sets that included students' math and verbal achievements and self-concepts. These effects generalized across different countries, gender distributions, and age groups. However, more recent studies examining the I/EM with elementary school students did not consistently find significantly negative dimensional comparison effects (e.g., Ehm, Lindberg, & Hasselborn, 2014; Möller, Kuska, & Zaunbauer, 2011; Weidinger, Steinmayr, & Spinath, 2019; but see Pinxten et al., 2015; Sewasew & Schroeders, 2019). This indicates that students start to contrast their math and verbal achievements only after some years at school. Furthermore, Möller et al.'s analysis revealed stronger social comparison effects when students' achievements were measured with grades, rather than standardized test scores. Marsh et al. (2014) replicated this finding in a more

recent study examining the effect of achievement operationalization on the comparison effects in the I/EM within the same sample. In this study, they additionally found stronger dimensional comparison effects where grades were used as achievement indicators. These findings demonstrate that grades are the more salient source of feedback students receive about their academic accomplishments and that students thus consider their grades in particular in the process of self-concept formation (i.e., the achievement information that they have in mind, rather than their results from standardized tests, or even their “true” achievement, which are usually unknown to them).

The Reciprocal I/E Model

To overcome the specific limitations of the REM (i.e., its confinement to one subject) and the I/EM (i.e., its confinement to the effects of achievements on self-concepts), Möller et al. (2011) combined these two models into the RI/EM, which describes the reciprocal relations between math and verbal achievements and self-concepts (see Figure 3). Most of the assumptions of the RI/EM are derived from the REM and the I/EM (see also Niepel, Brunner, & Preckel, 2014). First, the RI/EM predicts positive effects of achievements on subsequent self-concepts within subjects, even after controlling for prior self-concepts (the skill-development effects in the REM and the social comparison effects in the I/EM; α). Second, it predicts positive effects of self-concepts on subsequent achievements within subjects, even after controlling for prior achievements (the self-enhancement effects in the REM; β). Third, it assumes positive effects of achievements on subsequent achievements, as well as of self-concepts on subsequent self-concepts, within subjects (the autoregressive effects in the REM; ϵ). Fourth, it assumes positive effects of achievements on subsequent achievements between subjects (the longitudinal extension of the positive achievement correlation in the I/EM; ζ). Fifth, it assumes near-zero effects of self-concepts on subsequent self-concepts between subjects (the longitudinal extension of the near-zero self-concept correlation in the I/EM; η). Sixth, it predicts negative effects of achievements on subsequent self-concepts between

subjects (the dimensional comparison effects in the I/EM; γ). Finally, it postulates negative effects of self-concepts on subsequent achievements between subjects (δ). This prediction is not found explicitly in the REM, nor in the I/EM, but follows from adopting the logic of the I/EM to the reciprocal effects of self-concepts on achievements. As Möller et al. argue, students might perform dimensional comparisons not only between their math and verbal achievements but also between their math and verbal self-concepts (e.g., “I think that I’m more talented in math than in verbal subjects”). On grounds of an increased (decreased) motivation to deal with the subject with their intra-individually higher (lower) self-concept (e.g., Arens, Becker, & Möller, 2017; Schurtz, Pfoest, Nagengast, & Artelt, 2014), they might show a higher (lower) engagement and time investment and thus enhance (worsen) their achievement in this subject.

Empirically, the RI/EM has been tested in several studies (e.g., Chen, Yeh, Hwang, & Lin, 2013; Köller, Klemmert, Möller, & Baumert, 1999; Marsh & Köller, 2004; Marsh & Yeung, 1998; Möller & Köller, 2001; Möller et al., 2011; Möller, Zimmermann, & Köller, 2014; Niepel et al., 2014; Sewasew & Schroeders, 2019; Weidinger et al., 2019), although not all of them have used this label. Overall, these studies (especially those with secondary school students) have provided strong support for the model’s assumptions. In line with the REM, there were positive reciprocal effects within subjects, which tended to be stronger on self-concepts than on achievements, as well as positive autoregressive effects. In accord with the I/EM, the effects of achievements on achievements between subjects were positive, whereas most of the effects of self-concepts on self-concepts between subjects were close to zero (although these effects were only examined in approximately every second study). The effects of achievements on self-concepts between subjects were usually significantly negative. In contrast, the findings concerning the effects of self-concepts on achievements between subjects were rather mixed. Although these effects often tended to be negative, they were usually only small and nonsignificant, and sometimes even tended towards positive. Thus, it

is still an open question as to how students' self-concept in one subject impacts on their achievement in another.

The 2I/E Model

On the basis of empirical findings showing that students compare their achievements not only socially and dimensionally, but also temporally, when forming their self-concepts (e.g., Müller-Kalthoff et al., 2017; Wolff, Helm, Zimmermann, et al., 2018), Wolff and colleagues (Wolff, Helm, & Möller, 2019; Wolff, Nagy, et al., 2019) developed the 2I/EM, which takes this second internal comparison into account. Like the I/EM, the 2I/EM only considers the effects of achievements on self-concepts. However, it describes these effects in a more differentiated way by considering achievement information from different points in time, and separating students' achievement levels (i.e., their average achievement during the period considered; "How good am I in comparison to my classmates?") from their achievement changes (i.e., their development in achievement during this period; "How good am I in comparison to former times?") in math and verbal subjects (see Figure 4). In line with the I/EM, the 2I/EM assumes positive effects of achievement levels on self-concepts within subjects (social comparison effects; α) and negative effects of achievement levels on self-concepts between subjects (dimensional comparison effects; γ). Moreover, it assumes positive effects of achievement changes on self-concepts within subjects (temporal comparison effects; θ), indicating that achievement improvement (decline) leads to a higher (lower) self-concept. In addition, most studies of the 2I/EM have examined the effects of achievement changes on self-concepts between subjects (ι), interpreted as effects of dimensional comparisons between achievement changes, but assumed to be close to zero because of the complexity of this comparison process.

To date, the 2I/EM assumptions have found support in 11 empirical studies (Wolff, Helm, & Möller, 2019; Wolff, Helm, Zimmermann, et al., 2018; Wolff, Nagy, et al., 2019; Wolff, Wigfield, Möller, Dicke, & Eccles, 2019). Usually, these studies have revealed strong

positive social comparison effects, moderate negative dimensional comparison effects, and small positive temporal comparison effects across student samples of different age groups, school types, and countries. The effects of achievement changes on self-concepts between subjects were usually nonsignificant and close to zero. Wolff, Nagy, et al. (2019, Studies 1 and 2) compared the 2I/EM effects as a function of achievement operationalization. In line with studies testing the I/EM, they found stronger comparison effects when using grades rather than standardized test scores as achievement indicators.

A few studies have tested the 2I/EM while controlling for students' prior self-concepts, measured before the first achievement measure (Wolff, Wigfield, et al., 2019; Wolff, Nagy, et al., 2019, Study 4; see also Figure 4). In accord with the RI/EM, these studies assumed and found support for positive effects of self-concepts on self-concepts within subjects (ϵ), and near-zero effects of self-concepts on self-concepts between subjects (η). Furthermore, the social, dimensional, and temporal comparison effects remained significant, although the inclusion of students' prior self-concepts led to changes in the strength of these effects: Whereas the social and dimensional comparison effects decreased (which can be explained by the relatively high stability of students' achievement levels and self-concepts across time, which explains why students' achievement levels are already manifested in part in their prior self-concepts, which show positive autoregressive effects on their subsequent self-concepts), the temporal comparison effects usually increased, implying that achievement changes have a stronger impact on students' self-concept changes than on their self-concepts measured at a certain point in time (see Wolff, Wigfield, et al., 2019, for a more detailed discussion). The inclusion of students' prior self-concepts into the 2I/EM can be seen as a first step towards the development of the R2I/EM (because in the REM and RI/EM, effects of achievements on self-concepts are controlled for prior self-concepts also). However, the 2I/EM in its current form does not allow examining how students' achievements are affected by their self-concept levels and changes.

The Reciprocal 2I/E Model

In this article, we extend the 2I/EM into the R2I/EM, to examine not only the effects of students' math and verbal achievement levels and changes on their math and verbal self-concepts, but also the effects of students' math and verbal self-concept levels and changes on their math and verbal achievements. Thus, our research contributes to the open question of how self-concept relates to subsequent achievement within and especially between subjects, while for the first time examining the effects of self-concept changes on achievements. Our approach is inspired by the extension of the I/EM into the RI/EM. However, the RI/EM and the R2I/EM differ from each other conceptually: Whereas the RI/EM integrates alternately measured achievement and self-concept variables into one model, this is not possible in the R2I/EM, because of the double function of these variables as both indicators of the level and change factors and as criteria predicted by the other construct's level and change factors (or as control variables not used to form level and change factors). To solve this dilemma, the R2I/EM is composed of two sub-models, which show a large overlap in the included achievement and self-concept variables and thus allow juxtaposing the effects of achievements on self-concepts versus self-concepts on achievements during the same period. One sub-model corresponds to the 2I/EM, while controlling for prior self-concepts, as depicted in Figure 4. In the other sub-model, the achievement and self-concept variables are simply exchanged. Figure 5 shows this sub-model, which we designate the reverse 2I/EM (r2I/EM; where the lowercased r also expresses that the r2I/EM is part of the R2I/EM).

The r2I/EM assumptions are largely derived from the RI/EM: First, the model assumes positive effects of achievements on subsequent achievements within subjects (autoregressive effects; ϵ). Second, it assumes positive effects of achievements on subsequent achievements between subjects (longitudinal extension of the positive achievement correlation; ζ). Third, it predicts positive effects of self-concept levels on subsequent achievements (self-enhancement effects; β). Fourth, it predicts negative effects of self-concept levels on subsequent

achievements between subjects (dimensional self-concept comparison effects; δ). Fifth, it postulates positive effects of self-concept changes on subsequent achievements within subjects (κ), which result from adopting the logic of temporal comparison effects in the 2I/EM to the r2I/EM. In line with the argument of the RI/EM that students compare their math and verbal self-concepts dimensionally, we argue that they might also perform temporal comparisons of their math and verbal self-concepts (i.e., they note that they judge their subject-specific abilities as higher, lower, or similar compared to the past). As increases (decreases) in self-concept should enhance (lower) students' motivation to engage in subject-relevant activities, this should lead to increases (decreases) in students' achievements as a result of more (less) time spent for learning and practice. Finally, the r2I/EM includes effects of self-concept changes on subsequent achievements between subjects (λ), which can be interpreted as the effects of dimensional comparisons between self-concept changes. Similarly to the 2I/EM, however, we deem it likely that these effects would be rather close to zero, given the complexity of this comparison process, as well as the fact that the effects of self-concept (level) on achievement in RI/EM studies, if negative at all, were only small.

The Present Research

In this research, we tested the newly developed R2I/EM in three empirical studies. In Study 1, we examined a sample of secondary school students between Grade 4 and Grade 8 from Germany. In Study 2, we examined a sample of secondary school students between Grade 9 and Grade 11 from Switzerland. In Study 3, our sample consisted of elementary school students between Grade 4 and Grade 6 from Belgium. Whereas Studies 1 and 2 used grades as achievement indicators, in Study 3 students' achievement was measured with standardized tests. Study 1 and Study 2 included six measurement points, where students' math and verbal achievements (T1, T3, T5) and self-concepts (T2, T4, T6) were measured alternately. This design allowed us to test a 2I/EM in which students' self-concepts at T6 were predicted by their achievement levels and changes between T3 and T5, controlled for their

self-concepts at T2, as well as an r2I/EM, in which students' achievements at T5 were predicted by their self-concept levels and changes between T2 and T4, controlled for their achievements at T1. In Study 3, students' math and verbal achievements and self-concepts were measured at the same time points (T1, T2, T3). However, as the achievement tests covered the content taught during the past months, whereas the self-concept measures referred to students' beliefs at the respective points in time, we treated this design similarly to that of Study 1 and 2. Thus, we tested a 2I/EM in which students' self-concepts at T3 were predicted by their achievement levels and changes between T2 and T3, controlled for their self-concepts at T1, as well as an r2I/EM, in which students' achievements at T3 were predicted by their self-concept levels and changes between T1 and T2, controlled for their achievements at T1.

We hypothesized finding support for the R2I/EM assumptions in all three studies. In the 2I/EM sub-model, we expected to find (1) positive effects of self-concepts on subsequent self-concepts within subjects, (2) near-zero effects (i.e., $|\beta| < .10$) of self-concepts on subsequent self-concepts between subjects, (3) positive effects of achievement levels on self-concepts within subjects, (4) negative effects of achievement levels on self-concepts between subjects, (5) positive effects of achievement changes on self-concepts within subjects, and (6) near-zero effects (i.e., $|\beta| < .10$) of achievement changes on self-concepts between subjects. In the r2I/EM sub-model, we expected to find (7) positive effects of achievements on subsequent achievements within subjects, (8) positive effects of achievements on subsequent achievements between subjects, (9) positive effects of self-concept levels on achievements within subjects, (10) negative effects of self-concept levels on achievements between subjects, (11) positive effects of self-concept changes on achievements within subjects, and (12) near-zero effects (i.e., $|\beta| < .10$) of self-concept changes on achievements between subjects.

Study 1: Method

Sample

The sample stemmed from the larger project BLIND. It comprised $N = 1,508$ students of one cohort between Grade 4 and Grade 8 attending 60 secondary schools representative of the German federal state of BLIND (age at the end of Grade 4: $M = 10.9$, $SD = 0.56$; gender: 49.1% female, 50.9% male; migration background: 78.5% with both parents born in Germany, 15.1% with at least one parent born outside Germany, 6.4% without information about parents' origin). Students reported their math and German grades from their report cards at the end of Grade 4 (T1), after the first semester of Grade 6 (T3), and at the end of Grade 7 (T5). They reported their math and German self-concepts at the beginning of Grade 5 (T2), at the beginning of the second semester of Grade 6 (T4), and at the beginning of Grade 8 (T6). Data collection took place between 2004 and 2007 during regular school lessons. Providing grades and self-concepts was voluntary, and the informed consent of parents was required. The BLIND data set had already been used in previous research to examine the RI/EM (BLIND) and the 2I/EM (BLIND). In this way, the present analysis is not fully independent of these prior studies. However, the prior studies did not investigate the effects of self-concept levels and changes on achievement as postulated in the R2I/EM, which are of particular interest for the present research. Accordingly, the present study goes far beyond the prior ones by using the BLIND data for the first time to test the r2I/EM sub-model. In addition, the analysis testing the 2I/EM sub-model shows some improvements compared to the prior study, resulting in a better model fit (specification of latent self-concept variables by using the effects-coding method, no exclusion of students due to missing values, consideration of additional sociodemographic background variables; see Statistical Analyses section for more details).

Measures

Achievements. Students reported their math and German grades from their report cards. The German grading system ranges from 1 = *excellent* to 6 = *insufficient*. Therefore, we reverse-coded students' grades such that higher scores indicated better achievements.

Previous studies have shown that self-reports of grades in the German school system are sufficiently reliable (Dickhäuser & Plenter, 2005; Schneider & Sparfeldt, 2016). Thus, the number of students who did not report their actual grades should have been rather low. Notwithstanding this, we note that, in some respects, self-reported grades may be even more appropriate for the investigation of comparison effects than grades reported by schools, provided that some students do not remember their grades correctly and instead report and compare those grades they have in mind.

Self-concepts. Students' math and German self-concepts were assessed with three items worded in parallel, which originated in the Self Description Questionnaire II (Marsh, 1992b): (1) "I get good grades in math/German", (2) "Math/German is one of my best subjects", (3) "I have always done well in math/German". Students responded to all items on a 4-point Likert scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. Reliabilities were good (math: all $\omega = .94$; German: $.91 \leq \omega \leq .94$).

Statistical Analyses

We calculated the R2I/EM by means of structural equation modeling using Mplus 7.4 (Muthén & Muthén, 2015). For model estimation, we used the robust maximum likelihood (MLR) estimator implemented in Mplus, which is robust against violations of normality assumptions (Yuan & Bentler, 2000). We used the complex modeling procedure in Mplus (*type = complex*) to correct the estimated standard errors for the nestedness in the data (Hox, 2002). As no information about students' class affiliation was available, we used students' school as cluster variable.

We specified students' math and verbal self-concepts as latent constructs by using the effects-coding method (Little, Slegers, & Card, 2006) under scalar measurement invariance (Little, Preacher, Selig, & Card, 2007). Thus, we fixed the sum of the factor loadings of each self-concept construct to the number of the construct-specific indicators (i.e., 3) and the sum of the intercepts of all indicators of one construct to 0. Furthermore, we set the factor loadings

and the intercepts of the same self-concept indicators across time to be equal. Additional analyses showed that the assumption of scalar invariance was justified (see Table 1). We also allowed correlations between the same self-concept items across time, as well as between analogous self-concept items at one measurement point (Marsh, 1990b). Whereas self-concepts are psychological constructs which exist in students' minds and are not fully represented by the items of a self-concept scale (and thus require latent modeling), we assumed that students would refer to their received specific achievement feedback when comparing their achievements (insofar as they remember this feedback). Accordingly, we specified students' achievements as manifest variables. We did so in both sub-models of the R2I/EM for reasons of consistency and for the comparability of the effects between both sub-models. The approach of specifying students' self-concepts as latent variables and achievements as manifest variables was in line with prior studies of the RI/EM (Möller et al., 2011; Möller et al., 2014; Niepel et al., 2014) and the 2I/EM (Wolff, Helm, & Möller, 2019; Wolff, Helm, Zimmermann, et al., 2018; Wolff, Nagy, et al., 2019).

We defined the achievement and self-concept levels and changes for each subject according to the growth curve modeling approach (Meredith & Tisak, 1990). In 2I/EM, we specified the achievement levels by fixing the loadings of the achievement variables from T3 and T5 to 1, and the achievement changes by fixing the loadings of the achievement variables from T3 to -0.5 and from T5 to 0.5. To keep the model identified, we fixed the intercepts and residual variances of all achievement variables to 0. This approach was similar to prior studies of the 2I/EM (Wolff, Helm, & Möller, 2019; Wolff, Helm, Zimmermann, et al., 2018; Wolff, Nagy, et al., 2019). Our specification in the r2I/EM was analogous to that of the 2I/EM: We specified the self-concept levels by fixing the loadings of the latent self-concept variables from T2 and T4 to 1, and the self-concept changes by fixing the loadings of the latent self-concept variables from T2 to -0.5 and from T4 to 0.5. We fixed the intercepts and residual variances of all latent self-concept variables to 0, to keep the model identified. In the 2I/EM,

students' math and verbal self-concepts at T6 were predicted by their math and verbal achievement levels, achievement changes, and self-concepts at T2. In the r2I/EM, math and verbal achievements at T5 were predicted by math and verbal self-concept levels, self-concept changes, and achievements at T1. In both sub-models, we allowed correlations between all predictor variables, as well as between the residuals of the criteria. Before calculating the R2I/EM we conducted preliminary analyses including the achievement and self-concept variables from all measurement points without specifying level and change variables. In these analyses, we compared statistical parameters (e.g., means) with each other by using the *Constraint* option implemented in MPlus, which is based on the delta method (e.g., Raykov & Marcoulides, 2009).

On average, 21.3% of the values on the achievement variables (math: 12.9% at T1, 23.0% at T2, 27.7% at T3; German: 13.5% at T1, 22.7% at T2, 28.1% at T3) and 17.5% of the values on the self-concept variables (math: 12.2% at T1, 15.0% at T2, 25.2% at T3; German: 12.1% at T1, 15.3% at T2, 25.6% at T3) were missing. Most of these missing values were due to the absence of students during data collection. To handle missing values, we used the full information maximum likelihood (FIML) procedure, which is unbiased under the missing at random (MAR) assumption and retains statistical power as no observations are deleted (Enders, 2010). According to Marsh, Pekrun, et al. (2019), serious violations of the MAR assumption are unlikely in longitudinal studies. Even though missing values under the MAR assumption do not depend on the missing values themselves, they are allowed to depend on values of all other variables included in the analysis. Accordingly, missing values in longitudinal designs may be conditional on the values of the same variable collected at another measurement point. Because of the longitudinal design of our study, and because we considered students' achievements and self-concepts to be particularly predictive for missing values, we considered the MAR assumption as tenable. Nevertheless, we additionally

included sociodemographic background variables (age, gender, migration background) in our analysis to further improve the accuracy of the FIML algorithm.

We judged the fit of the models using the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA). Usually, CFI and TLI values greater than .95 (.90) and RMSEA values lower than .05 (.08) are interpreted as reflecting a good (acceptable) fit to the data (e.g., Hu & Bentler, 1999; Schweizer, 2010; West, Taylor, & Wu, 2012). We referred to the conventional significance level of $\alpha = .05$ to test our hypotheses assuming positive or negative effects. We used equivalence tests to test our hypotheses assuming near-zero effects by testing against boundary values of $\beta = -.10$ and $\beta = .10$ (e.g., Lakens, Scheel, & Isager, 2018). Our syntaxes for calculating the R2I/EM, as well as the data used in all three studies of this research, can be found via the following link: <https://osf.io/7f6gb/> (BLIND).

Study 1: Results

Preliminary Analyses

Table 2 shows the means, standard deviations, and correlations of the model, including the achievement and self-concept variables from all measurement points. It demonstrates that students' achievements decreased between T1 and T3 (math: $\Delta M = -0.30$, $p < .001$, $d = 0.28$; German: $\Delta M = -0.31$, $p < .001$, $d = 0.32$), but remained stable between T3 and T5 (math: $\Delta M = -0.04$, $p = .40$; German: $\Delta M = -0.00$, $p = .96$). Similarly, students' German self-concept decreased between T2 and T4 ($\Delta M = -0.17$, $p < .001$, $d = 0.23$), while remaining stable between T4 and T6 ($\Delta M = -0.01$, $p = .71$). Students' math self-concept decreased both between T2 and T4 ($\Delta M = -0.18$, $p < .001$, $d = 0.23$) and between T4 and T6 ($\Delta M = -0.06$, $p = .05$, $d = 0.09$). Although students' math and German achievements were similar at each measurement point (all $\Delta M \leq 0.06$, all $p \geq .10$), their self-concepts were higher in math than in German (all $0.21 \leq \Delta M \leq 0.27$, all $p < .001$, all $0.19 \leq d \leq 0.26$). Students' subject-specific achievements showed moderate to high correlations across time (all $.31 \leq r \leq .50$, all $p <$

.001). The correlations between their subject-specific self-concepts across time were even somewhat higher (all $.43 \leq r \leq .66$, all $p < .001$). Whereas students' math and German achievements showed strong positive correlations at each measurement point (all $.46 \leq r \leq .55$, all $p < .001$), the correlations between their math and German self-concepts were small and positive at T2 ($r = .13$, $p < .001$) and close to zero at T4 and T6 (all $.04 \leq r \leq .06$, all $p \geq .11$).

Testing the Reciprocal 2I/E Model

Table 3 presents the results of the R2I/EM. Moreover, it shows that the fit of both sub-models was satisfactory (all CFI $\geq .964$, all TLI $\geq .944$, all RMSEA $\leq .057$). In the 2I/EM sub-model, all six hypotheses found support. There were moderate positive effects of self-concepts on subsequent self-concepts within subjects (all $.21 \leq \beta \leq .25$, all $p < .001$) and near-zero effects of self-concepts on subsequent self-concepts between subjects, which were significantly different from the boundary values (all $\beta = -.04$, all $p \geq .20$, all 90% CI $[-.097, .016]$). The effects of achievement levels on self-concepts were strongly positive within subjects (all $.63 \leq \beta \leq .66$, all $p < .001$) and moderately negative between subjects (all $-.22 \leq \beta \leq -.20$, all $p < .001$). Furthermore, there were small to moderate positive effects of achievement changes on self-concepts within subjects (all $.16 \leq \beta \leq .21$, all $p < .001$), whereas the effects of achievement changes on self-concepts between subjects were close to zero and significantly different from the boundary values (all $-.02 \leq \beta \leq .02$, all $p \geq .37$, all 90% CI $[-.061, .056]$). In the r2I/EM sub-model, our hypotheses also found strong support: In line with our assumptions, there were small positive effects of achievements on subsequent achievements within subjects (all $\beta = .14$, all $p \leq .02$). Furthermore, we found a small positive effect of German achievement on subsequent math achievement ($\beta = .16$, $p < .01$). The effect of math achievement on subsequent German achievement tended to be positive but failed to reach statistical significance ($\beta = .09$, $p = .08$). As expected, we found moderate positive effects of self-concept levels on achievements within subjects (all $\beta = .36$, all $p < .001$) and

small positive effects of self-concept changes on achievements within subjects (all $.17 \leq \beta \leq .20$, all $p < .001$). In contrast to our assumptions, however, the effects of self-concept levels on achievements between subjects were close to zero and nonsignificant, rather than negative (all $-.01 \leq \beta \leq .03$, all $p \geq .36$). Moreover, the effect of math self-concept change on German achievement was significantly positive, rather than close to zero ($\beta = .10$, $p < .01$, 90% CI [.047, .160]). The effect of German self-concept change on math self-concept was nonsignificant but also tended to be positive and did not significantly differ from the upper boundary value ($\beta = .08$, $p = .08$, 90% CI [.006, .146]).

Study 2: Method

Sample

The sample stemmed from the larger project BLIND. It comprised $N = 748$ students of one cohort between Grade 9 and Grade 11 attending eight academic track secondary schools in the German-speaking part of Switzerland (age in the middle of the second semester of Grade 9: $M = 15.6$, $SD = 0.60$; gender: 56.4% female, 43.6% male). All students had German as their native language (migration background: 92.0% born in Switzerland, 7.2% not born in Switzerland, 0.8% without information about native country). Students' math and German grades were derived from their report cards after the first semester of Grade 9 (T1), Grade 10 (T3), and Grade 11 (T5). Students reported their math and German self-concepts in the middle of the second semester of Grade 9 (T2), Grade 10 (T4), and Grade 11 (T6), about four months after they had received their report cards. Data collection took place between 2012 and 2014 during regular school lessons. Participation was voluntary and the informed consent of parents was required.

Measures

Achievements. Students' math and German grades from their report cards were provided by school administrators. In Switzerland, grades range from 1 = *insufficient* to 6 = *excellent*, including half grades (e.g., 5.5). Thus, higher scores indicated better achievements.

Self-concepts. Students' math and German self-concepts were assessed with the same three items as in Study 1. Students responded to all items on a 5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*. Reliabilities were satisfactory (math: $.88 \leq \omega \leq .89$; German: all $\omega = .90$).

Statistical Analyses

The statistical analyses were similar to Study 1, except for the fact that we used students' class rather than their school as cluster variable. The average number of missing values was 21.6% on the achievement variables (math: 3.5% at T1, 26.1% at T2, 35.7% at T3; German: 3.2% at T1, 25.3% at T2, 35.7% at T3) and 24.3% on the self-concept variables (math: 2.1% at T1, 32.4% at T2, 38.4% at T3; German: 2.2% at T1, 32.3% at T2, 38.5% at T3). A significant number of missing values was due to the dropout of one school after T1 ($n = 123$).

Study 2: Results

Preliminary Analyses

Table 4 shows the means, standard deviations, and correlations of the model, including the achievement and self-concept variables from all measurement points. It demonstrates that students' math achievement decreased between T1 and T3 ($\Delta M = -0.15$, $p < .01$, $d = 0.21$) and remained stable between T3 and T5 ($\Delta M = -0.02$, $p = .64$), whereas students' German achievement remained stable between T1 and T3 ($\Delta M = -0.03$, $p = .33$), but slightly increased between T3 and T5 ($\Delta M = 0.07$, $p = .03$, $d = 0.14$). Students' math and German self-concepts decreased between T2 and T4 (math: $\Delta M = -0.19$, $p < .001$, $d = 0.26$; German: $\Delta M = -0.11$, $p < .001$, $d = 0.16$), while remaining stable between T4 and T6 (math: $\Delta M = -0.02$, $p = .69$; German: $\Delta M = 0.02$, $p = .52$). At all measurement points, students showed higher German than math achievements (all $0.15 \leq \Delta M \leq 0.35$, all $p < .01$, all $0.19 \leq d \leq 0.42$) and self-concepts (all $0.20 \leq \Delta M \leq 0.32$, all $p \leq .02$, all $0.14 \leq d \leq 0.21$). Students' subject-specific achievements showed high correlations across time (all $.49 \leq r \leq .68$, all $p < .001$). The

correlations between their subject-specific self-concepts across time were even somewhat higher (all $.65 \leq r \leq .79$, all $p < .001$). Whereas students' math and German achievements showed positive correlations at each measurement point (all $.22 \leq r \leq .29$, all $p < .001$), the correlations between their math and German self-concepts were or tended to be negative (all $-.26 \leq r \leq -.09$, all $p \leq .06$).

Testing the Reciprocal 2I/E Model

Table 5 presents the results of the R2I/EM. Moreover, it shows that the fit of both sub-models was satisfactory (all CFI $\geq .958$, all TLI $\geq .934$, all RMSEA $\leq .064$). In the 2I/EM sub-model, all six hypotheses found support, although our second hypothesis only partly. There were strong positive effects of self-concepts on subsequent self-concepts within subjects (all $.41 \leq \beta \leq .44$, all $p < .001$) and near-zero effects of self-concepts on subsequent self-concepts between subjects (all $.00 \leq \beta \leq .04$, all $p \geq .36$). However, whereas the effect of German self-concept on math self-concept did not significantly differ from the boundary values (90% CI $[-.057, .065]$), the effect of math self-concept on German self-concept did (90% CI $[-.033, .115]$). As expected, the effects of achievement levels on self-concepts were strongly positive within subjects (all $.50 \leq \beta \leq .59$, all $p < .001$) and small to moderately negative between subjects (all $-.30 \leq \beta \leq -.13$, all $p \leq .01$). Furthermore, there were small positive effects of achievement changes on self-concepts within subjects (all $\beta = .16$, all $p < .001$), whereas the effects of achievement changes on self-concepts between subjects were close to zero and significantly different from the boundary values (all $-.02 \leq \beta \leq .03$, all $p \geq .49$, all 90% CI $[-.059, .093]$). In the r2I/EM sub-model, our hypotheses also found support, with one exception: In line with our assumptions, there were strong positive effects of achievements on subsequent achievements within subjects (all $.39 \leq \beta \leq .42$, all $p < .001$) and small to moderate positive effects of self-concepts on subsequent self-concepts between subjects (all $.16 \leq \beta \leq .32$, all $p = .02$). The effects of self-concept levels on achievements were small to moderately positive within subjects (all $.15 \leq \beta \leq .29$, all $p \leq .01$), and small to

moderately negative between subjects (all $-.26 \leq \beta \leq -.14$, all $p < .01$). Moreover, we found small to moderate positive effects of self-concept changes on achievements within subjects (all $.18 \leq \beta \leq .22$, all $p < .01$). However, the effects of self-concept changes on achievements between subjects did not significantly differ from the upper boundary value. Whereas the effect of German self-concept change on math achievement was nonsignificant ($\beta = .06$, $p = .17$, 90% CI $[-.012, .133]$), the effect of math self-concept change on German achievement was even significantly positive ($\beta = .13$, $p < .01$, 90% CI $[.095, .274]$).

Study 3: Method

Sample

The sample stemmed from the larger project BLIND. It consisted of $N = 3,710$ students of one cohort between Grade 4 and Grade 6 attending 120 elementary schools from Flanders, the Dutch-speaking part of Belgium (age at the end of Grade 4: $M = 10.04$, $SD = 0.49$; gender: 49.9% female, 50.1% male; migration background: 78.4% born in Belgium, 4.0% not born in Belgium, 17.5% without information about country of birth). These schools were representative of the total of all elementary schools in Flanders. Students' math and Dutch achievements and self-concepts were assessed between 2007 and 2009 at three measurement points, at the end of Grade 4 (T1), Grade 5 (T2), and Grade 6 (T3), during regular school lessons. Schools were provided with feedback about their individual students' achievements. Thus, it seemed likely that students were aware of their performance in the achievement tests and compared their results from the different tests. Participation in the research project was voluntary and the informed consent of parents was required. We note that the BLIND data have already been analyzed on one prior I/EM study (BLIND). However, this study was cross-sectional and only used the data from Grade 4.

Measures

Achievements. Students' achievements were measured with items from standardized tests. The math achievement tests covered five broad areas: mental arithmetic, arithmetic,

algebraic problems, geometry, and applied arithmetic. The Dutch achievement tests consisted of reading comprehension tasks. Each test was available in two versions, of varying difficulty, and students were assigned to one version on the basis of their achievements in the previous school year. All scales were formed within the context of item-response theory, with overlapping items, by computing weighted likelihood estimates. The reliabilities of the tests were satisfactory (math: $.88 \leq \alpha \leq .90$; Dutch: $.76 \leq \alpha \leq .86$). The correlations between the subject-specific test scores were high (math: $.81 \leq r \leq .86$; Dutch: $.77 \leq r \leq .83$).

Self-concepts. Students' math and Dutch self-concepts were assessed with three items worded in parallel, which originated from the Self Description Questionnaire I (Marsh, 1992a): (1) "I get good grades in math/Dutch", (2) "I am good at math/Dutch", (3) "I am learning fast in math/Dutch". Students responded to all items on a 5-point Likert scale ranging from 1 = *not true* to 5 = *true*. Reliabilities were satisfactory (math: $.85 \leq \omega \leq .92$; Dutch: $.86 \leq \omega \leq .92$).

Statistical Analyses

The statistical analyses were similar to Study 1, except for the fact that we used students' achievements from T2 and T3 to form their achievement levels and changes as predictors of their self-concepts at T3, while controlling for their self-concepts from T1 (2I/EM), and students' self-concepts from T1 and T2 to form their self-concept levels and changes as predictors of their achievements at T3, while controlling for their achievements from T1 (r2I/EM). The average number of missing values was 11.0% on the achievement variables (math: 11.9% at T1, 10.5% at T2, 10.3% at T3; Dutch: 12.0% at T1, 10.8% at T2, 10.4% at T3) and 13.5% on the self-concept variables (math: 16.0% at T1, 12.7% at T2, 11.9% at T3; Dutch: 15.7% at T1, 12.5% at T2, 12.0% at T3).

Study 3: Results

Preliminary Analyses

Table 6 shows the means, standard deviations, and correlations of the model, including the achievement and self-concept variables from all measurement points. It demonstrates that students' achievements increased from T1 to T2 (math: $\Delta M = 5.48, p < .001, d = 1.25$; Dutch: $\Delta M = 4.08, p < .001, d = 0.89$) and from T2 to T3 (math: $\Delta M = 3.64, p < .001, d = 0.81$; Dutch: $\Delta M = 5.18, p < .001, d = 1.20$). Students' math self-concepts decreased slightly from T1 to T2 ($\Delta M = -0.15, p < .001, d = 0.22$), but remained stable from T2 to T3 ($\Delta M = 0.01, p = .77$). Their Dutch self-concepts also decreased from T1 to T2 ($\Delta M = -0.16, p < .001, d = 0.21$), but increased from T2 to T3 ($\Delta M = 0.14, p < .001, d = 0.18$). Students' subject-specific achievements showed high correlations across time (all $.77 \leq r \leq .86$, all $p < .001$). The correlations between their subject-specific self-concepts were also strongly positive (all $.53 \leq r \leq .77$, all $p < .001$). At each measurement point, the correlations between students' math and Dutch achievements were strongly positive (all $.59 \leq r \leq .64$, all $p \leq .001$). The correlations between their math and Dutch self-concepts were small and positive (all $.07 \leq r \leq .11$, all $p < .01$). Students' math achievements showed strong positive correlations with their math self-concepts at each measurement point (all $.54 \leq r \leq .64$, all $p < .001$). In contrast, the correlations between students' Dutch achievements and self-concepts were only moderately positive (all $.25 \leq r \leq .33$, all $p \leq .001$).

Testing the Reciprocal 2I/E Model

Table 7 presents the results of the R2I/EM. Moreover, it shows that the fit of both sub-models was good (all CFI $\geq .989$, all TLI $\geq .982$, all RMSEA $\leq .032$). In both sub-models, our hypotheses found strong support. In the 2I/EM sub-model, we found strong positive effects of self-concepts on subsequent self-concepts within subjects (all $.42 \leq \beta \leq .46$, all $p < .001$) and near-zero effects of self-concepts on subsequent self-concepts between subjects, which were significantly different from the boundary values (math on Dutch: $\beta = -.01, p = .67$, 90% CI $[-.054, .032]$; Dutch on math: $\beta = -.06, p < .001$, 90% CI $[-.089, -.036]$). The effects of achievement levels on self-concepts within subjects were strongly positive for math ($\beta = .45$,

$p < .001$) and moderately positive for Dutch ($\beta = .29, p < .001$). The effects of achievement levels on self-concepts between subjects were slightly negative (all $-.11 \leq \beta \leq -.05$, all $p < .01$). The effects of achievement changes on self-concepts within subjects were slightly positive for math ($\beta = .06, p < .001$), but not for Dutch ($\beta = -.00, p = .98$). The effects of achievement changes on self-concepts between subjects were close to zero and significantly different from the boundary values (math on Dutch: $\beta = .04, p = .04$, 90% CI [.008, .074]; Dutch on math: $\beta = .03, p = .06$, 90% CI [.004, .052]). In the r2I/EM sub-model, we found strong positive effects of achievements on subsequent achievements within subjects (all $.60 \leq \beta \leq .61$, all $p < .001$) and small to moderate positive effects of achievements on subsequent achievements between subjects (all $.14 \leq \beta \leq .25$, all $p < .001$). The effects of self-concept levels on achievements were slightly positive within subjects (all $.10 \leq \beta \leq .18$, all $p \leq .001$), but close to zero (rather than negative) between subjects (all $-.02 \leq \beta \leq -.01$, all $p \geq .18$). The effects of self-concept changes on achievements within subjects were slightly positive for math ($\beta = .08, p < .001$), but not for Dutch ($\beta = .01, p = .77$). The effects of achievement changes on self-concepts between subjects were close to zero and did not significantly differ from the boundary values (math on Dutch: $\beta = .04, p = .03$, 90% CI [.009, .063]; Dutch on math: $\beta = -.01, p = .61$, 90% CI [-.030, .016]). Yet, it is worth noting that the effect of math self-concept change on Dutch achievement was significantly positive.

Discussion

The present research contributes substantially to our knowledge about the reciprocal relations between achievements and self-concepts in math and verbal subjects. In three empirical studies, we examined both the effects of achievement levels and changes on self-concepts and the effects of self-concept levels and changes on achievements in our newly developed R2I/EM. Our particular focus was on the effects of students' self-concept changes on their achievements, as these effects—to the best of our knowledge—have not been examined in any prior study. For the first time, we have been able to show that self-concept

changes, which we interpreted as temporal comparisons of self-concepts, affect students' achievements within and between subjects. In the following discussion, we will examine our findings in more detail. Subsequently, we point out implications, limitations, and directions for future research.

Testing the Reciprocal 2I/E Model

Overall, our three empirical studies provided strong support for the R2I/EM across student samples of different ages, school forms, and countries, as well as different forms of achievement operationalization (see Table 8 for an overview of the results in all studies). In the 2I/EM sub-model, we were able to replicate findings from previous research (e.g., Wolff, Helm, & Möller, 2019; Wolff, Nagy, et al., 2019) and thus to provide further support for the simultaneous operation of social, dimensional, and temporal comparisons in the process of self-concept formation. In Studies 1 and 2 (with grades as achievement indicators), we found strong positive social comparison effects, moderate negative dimensional comparison effects (except from a weaker dimensional comparison effect on students' math self-concept in Study 2), and small to moderate positive temporal comparison effects. In Study 3 (with test scores as achievement indicators), these effects were somewhat weaker but still statistically significant; this also corresponds to results from prior studies testing the 2I/EM with test scores (Wolff, Nagy, et al., 2019, Studies 1 and 2). Only the temporal comparison effect on students' Dutch self-concept was close to zero and nonsignificant. This finding could be due to the fact that the Dutch test in Study 3 focused on reading comprehension, and so only in part reflected students' abilities relevant to success in Dutch class—this assumption was further supported by the relatively small correlations between students' Dutch achievements and Dutch self-concepts, as well as the relatively small social comparison effect on students' Dutch self-concept. Thus, it is possible that a change in students' reading comprehension achievement did not have sufficient power to affect their Dutch self-concept.

The effects of achievement changes on self-concepts between subjects were close to zero in all studies. This further underpins the finding that students usually do not consider their achievement changes in one subject when forming their self-concept in another (Wolff, Helm, & Möller, 2019). Moreover, all three studies revealed moderate to strong positive autoregressive effects of self-concepts on subsequent self-concepts within subjects, but near-zero effects of self-concepts on subsequent self-concepts between subjects (even though the effect of math self-concept on German self-concept in Study 2, which had the smallest sample size, was not significantly weaker than $\beta = .10$). These findings also were in accord with prior studies of the 2I/EM (Wolff, Nagy, et al., 2019; Wolff, Wigfield, et al., 2019) and the RI/EM (e.g., Möller et al., 2011; Niepel et al., 2014). It is worth noting that the positive autoregressive effects within subjects were somewhat weaker in Study 1 than in Studies 2 and 3. One explanation for this might be that the first self-concept measurement in Study 1 took place immediately after students' transition from elementary school to secondary school, a period that often entails substantial changes in students' self-concepts (e.g., Arens & Watermann, 2015; Wigfield, Eccles, Iver, Reuman, & Midgley, 1991; Wolff, Wigfield, et al., 2019). Moreover, the interval between the two self-concept measures was somewhat longer in Study 1 than in Studies 2 and 3.

We were particularly interested in the relations of the r2I/EM sub-model, which we tested for the first time in the present research. Similarly to the 2I/EM, we found support for most of our hypotheses in the r2I/EM: In accord with previous research on the REM (e.g., Marsh & Martin, 2011; Marsh & O'Mara, 2008) and the RI/EM (e.g., Möller et al., 2011; Niepel et al., 2014), all three studies revealed positive effects of self-concept levels on achievements within subjects. These effects were small to moderate, and thus somewhat weaker than the effects of achievement levels on self-concepts within subjects. Thus, our results corroborate prior findings suggesting stronger skill-development than self-enhancement effects (e.g., Marsh & Yeung, 1997a; Möller et al., 2014).

Our findings concerning the effects of self-concept levels on achievements between subjects were mixed: Whereas these effects were small to moderately negative in Study 2, they were close to zero and nonsignificant in Studies 1 and 3. However, although it is not entirely in accord with our assumptions, this finding corresponds to prior findings from studies of the RI/EM, which likewise found no consistent evidence for the assumed negative effects of self-concepts on achievements between subjects (e.g., Möller et al., 2014; Sewasew & Schroeders, 2019). With respect to the present research, it is striking that the negative effects of self-concept levels on achievements between subjects were shown in the study with the older students. As suggested by the negative self-concept correlations, and in accordance with prior findings (e.g., Möller et al., 2009), these students seemed to have more differentiated self-concepts than those in the other samples. Hence, one could speculate that students start to compare their self-concept levels, and consequently shift their engagement to the subject with their intra-individually higher self-concept, only in higher grade levels. Moreover, it is worth noting that dimensional comparisons of self-concepts require a higher degree of introspection than dimensional comparisons of specific grades or test scores, and may therefore represent a more complex psychological process. This may be another reason as to why the negative effects of self-concept levels on achievements between subjects were only found in the sample including the older students.

Our investigation of the effects of self-concept changes on achievements yielded a particularly interesting picture: In line with our assumptions, all studies showed small to moderate positive effects of self-concept changes on achievements within subjects (although in Study 3 this effect was only shown on students' math self-concept, which again might be due to the restriction of the Dutch test to reading comprehension). These positive effects suggest that an increase (decrease) in self-concept enhances (decreases) students' engagement in the respective subject and thus leads to a better (worse) performance. However, all studies also revealed small positive effects of math self-concept change on verbal achievement

(although in Study 3 this effect was very small). Furthermore, the effects of verbal self-concept change on math achievement, although nonsignificant, mostly tended to be positive. This finding is especially remarkable as dimensional comparisons generally lead to negative effects between math and verbal subjects. Yet, this principle apparently does not apply to dimensional comparisons between self-concept changes. Rather, it seems that self-concept changes, especially in math, lead to assimilation effects on verbal achievement. Thus, a positive (negative) self-concept development in math might encourage (discourage) students to enhance their effort in both math and verbal subjects, rather than shifting their engagement towards math (verbal) subjects, eventually resulting in an overall positive (negative) development of math and verbal achievements. It remains unclear why, among the effects of self-concept changes on achievements between subjects, only those on verbal achievement became significant. Perhaps it is easier for students to impact on their verbal achievement, in contrast to their math achievement, by altering the amount of time they invest in these subjects.

Across all studies, we found the expected positive effects of achievements on subsequent achievements within and between subjects (although the effect of math achievement on German achievement in Study 1 failed to reach statistical significance). The strength of these effects ranged from small to strong within subjects and from small to moderate between subjects. The autoregressive effects were the strongest in Study 3. This finding makes sense, given that students' achievements in Study 3 were measured with standardized tests, which show less variation across time than do grades (because grades do not reflect merely objective achievements, but are also based on criteria such as engagement and effort). Similarly to the 2I/EM sub-model, we found the weakest autoregressive effects in the r2I/EM sub-model in Study 1. Again, this finding might be due to the longer interval between the two achievement measures in Study 1, as well as students' transition from

elementary school to secondary school. As grading standards differ between elementary and secondary school, a relatively low stability of students' achievements in Study 1 is plausible.

The Reciprocal 2I/E Model With Grades and Test Scores

With regard to the generalizability of R2I/EM we consider it an important advantage that we have found support for the model with both grades (reported by students and obtained by schools) and test scores as achievement indicators. Nevertheless, it is worth noting that the relations between achievements and self-concepts were usually stronger in the studies using grades to operationalize students' achievements. The stronger effects in the 2I/EM sub-model were in accord with prior studies (Wolff, Nagy, et al., 2019). Thus, our research provides further support for the assumption that students are more inclined to compare their grades from their report cards than their results from standardized tests. Reasons for this may be that students know their grades well (although in Study 3 students probably also got feedback about their performance in the achievement tests) and that grades are highly relevant to many students. The stronger effects on grades in the r2I/EM sub-model might be due to the fact that grades are not pure indicators of actual performance but also reflect behavior-related factors such as effort and diligence. Accordingly, the stronger effects on grades than on test scores corroborate our presumed explanation for the effects of self-concept levels and changes on achievements, that students adjust their subject-specific engagement as a function of their self-concept constellation, which subsequently impacts on their actual performance.

The operationalization of achievements through grades entails at least two idiosyncrasies that researchers should take into account when interpreting our findings, especially with regard to the analysis of achievement changes. First, teachers allocate grades according to performance standards that are specific to the respective grade level. Accordingly, a student is expected to show a higher actual performance (e.g., solving more difficult tasks, showing higher knowledge and skills) to get the same grade in a higher grade level. For example, this explains the seemingly paradoxical finding that students' average

achievements in Study 1 and 2 (measured through teachers-assigned grades) decreased in the course time, although students' actual competencies are very likely to have increased (just not to the desired extent)—whereas students' achievements in Study 3 (reflecting actual competencies) continually increased. Second, grades cannot change continuously in one direction because of the restriction of the grading scale. In contrast, this is possible if achievement is measured through standardized tests including suitable items. For example, if a student got the best grade in a subject, he or she cannot achieve an even better grade in this subject in the future. However, it should be noted that only a few students receive the best or worst possible grades in their report cards, at least in Germany and Switzerland. More specifically, in the present research less than 5% of all students in Study 1 and less than 3% of all students in Study 2 received these extreme grades. Hence, our findings should not have been significantly biased by such ceiling and floor effects. Notwithstanding, researchers should keep in mind the differences between grades and test scores as achievement indicators when comparing the results of the different studies of this research using either grades or test scores as achievement indicators.

Theoretical Implications

The present research significantly extends our knowledge about the generalizability of comparison processes, and especially temporal comparisons. Our findings suggest that students not only compare their subject-specific achievements, which results in impacts on their self-concepts, but also their subject-specific self-concepts—and that these comparisons have substantial effects on their achievements, presumably due to a changed engagement in subject-specific activities. However, the effects of achievement levels and changes on self-concepts do not always correspond to the respective effects of self-concept levels and changes on achievements. In particular, this is the case for the effects between subjects: Our studies consistently revealed negative effects of students' achievement levels on their self-concepts between subjects, implying that students enhance their self-concept in the subject with their

intra-individually higher achievement and worsen their self-concept in the subject with their intra-individually lower achievement. In contrast, the effects of students' self-concept levels on their achievements between subjects were often nonsignificant, especially for younger students, implying that students' perception of their abilities in one subject is of low importance for their achievement development in another subject. Furthermore, all of our studies showed near-zero effects of achievement changes on self-concepts between subjects, suggesting that students do not consider temporal comparisons of their achievements in one subject for assessing their self-concept in another. On the contrary, the effects of self-concept changes on achievements were significantly positive (on students' math achievement, although the effect on test scores was very low) or mostly tended to be positive (on students' verbal achievement). Thus, it seems that a change in students' subject-specific self-concepts, possibly via a change in general school involvement, also affects their achievements, especially their grades, in other subjects.

Although our empirical findings did not entirely support all assumptions of the R2I/EM, altogether, this model has proved itself as a valid and promising framework to examine the reciprocal effects between achievements and self-concepts. This is especially true in that most of the (few) deviations from our hypotheses, those related to the effects of self-concept levels and changes on achievements between subjects, occurred across multiple studies. However, given these countervailing results, we deem it indicated to slightly undermine the assumptions of the R2I/EM and propose that future research should assume either negative or near-zero effects of self-concept levels on achievements between subjects and either near-zero or positive effects of self-concept changes on achievements between subjects (see also, note in Figure 5).

Practical Implications

In our view, the most central new insight from this research is the fact that not only students' self-concept levels, but also their self-concept changes, seem to affect their

subsequent achievements. This finding has important educational implications, as it demonstrates that practitioners, such as teachers, can help students enhance their performance by impacting on their self-perceptions. It was already known, from studies of the RI/EM, that students' self-concept levels show positive effects on their subsequent achievements within subjects, whereas the effects of their self-concept levels on their subsequent achievements between subjects are often negative. Although this finding illustrates the significance of a high self-concept in one subject for further improvement in achievement in this subject, it also implies that a high self-concept in one subject can be counterproductive for students' development of achievement in another. In contrast, the finding of the present study that students' self-concept changes positively affect (or at least tend to affect) their subsequent achievements within and between subjects is quite promising, as it implies the general expediency for teachers to support their students' self-concepts. For example, this could be done by adopting an individualized teachers' frame of reference (e.g., Lüdtke, Köller, Marsh, & Trautwein, 2005), so that teachers pay particular attention to students' individual improvements when evaluating students' attainments, or by offering students attribution strategies (e.g., Weiner, 1986) that are likely to be beneficial for their self-concepts (e.g., attribution to talent and ability in the case of success, vs. attribution to external circumstances or the use of unfavorable learning strategies in the case of failure). Furthermore, teachers could make use of educational interventions explicitly aimed at enhancing students' self-concepts (see e.g., Craven, Marsh, & Burnett, 2004, for an overview).

Importantly, practitioners should bear in mind that the effects of self-concept changes on subsequent achievements have shown to be only small to moderate. Given that interventions to foster students' self-concepts can be completed in a relatively short time they can be considered as a rather efficient means to enhance students' achievements (e.g., compared to expensive private lessons). Nevertheless, the positive effects of self-concept increases on students' achievements are more likely to become noticeable in the long run,

after teachers have continually fostered their students' self-concepts. This is especially true for the effects within subjects, which in our research proved to be about twice as strong as the effects between subjects. Thus, teachers are advised to specifically promote their students' self-concepts in different subjects, rather than relying on assimilation effects, to achieve the highest possible achievement gains for their students through impacting on their self-concepts.

Another question that arises when thinking about self-concept enhancement as a way of promoting students' achievement is whether self-concepts can be increased without limit. This question goes beyond the scope of the present research. Notwithstanding, we would find it plausible that self-concepts, similarly to competencies, can increase continuously, although the effort to realize this increase is likely to intensify and the assessment of this increase might become more and more difficult at a certain point. For example, teachers should be able to also support students with extremely high self-concepts to believe even more strongly in their abilities by repeatedly pointing out the increasing demands that students have mastered during their school career. Notwithstanding, when promoting their students' self-concepts, teachers should ensure that these self-concepts do not become too unrealistically high, as students also benefit from realistic self-concepts, for example when making educational decisions such as course or career choices (e.g., Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Parker et al., 2012). Moreover, some researchers have warned that overly high self-concepts may lead to overconfidence that is associated with underachievement, although there is no clear empirical support for this assumption to date (see e.g., Trautwein & Möller, 2016, for a discussion). Given these deliberations, it generally seems advisable for teachers, if students already have high self-concepts, to support these students primarily in maintaining their self-concepts to counter achievement losses due to decreases in self-concepts, rather than relentlessly working on further self-concept enhancement. More generally, it seems useful to enlighten teachers about the complex interplay of achievements and self-concepts, so that they

can support their students successfully in the development of high achievements and high (but realistic) self-concepts.

Limitations and Directions for Future Research

Although our research provides substantial support for the validity and generalizability of the R2I/EM, it has some limitations that we consider worth discussing. One limitation is the fact that the R2I/EM only describes relations between achievements and self-concepts, while the mechanisms leading to these effects are at present just an interpretation. For the effects of achievements on self-concepts, previous studies have provided ample evidence for the assumption that these effects, at least in part, result from achievement comparisons. These studies have used various other methodological approaches, such as experimental (e.g., Müller-Kalthoff et al., 2017; Wolff, Helm, Zimmermann, et al., 2018), diary (e.g., Möller & Husemann, 2006; Olson & Evans, 1999), and vignette studies (e.g., Müller-Kalthoff et al., 2017; Wolff, Helm, & Möller, 2018). In contrast, it is still unclear in what way students compare their self-concepts and how these comparisons affect their learning activities (e.g., whether students really do show higher engagement in subject-specific activities in the case of self-concept enhancement). Future studies should address this question by including the mediating variables assumed to explain the effects of self-concepts on achievements in the R2I/EM. To this end, future studies could take advantage of ambulatory assessments examining students' self-concepts and subject-related activities in real-time and in students' everyday lives.

A second limitation of the R2I/EM is its confinement to the math and verbal domain. Previous research has shown that students compare their achievements not only in math and native language, but also in other school subjects (e.g., Arens, Helm, Wolff, & Möller, 2020; Jansen, Schroeders, Lüdtke, & Marsh, 2015) and in domains external to the school context (e.g., Tietjens, Möller, & Pohlmann, 2005; Wolff, Helm, Preckel, & Möller, 2018). These studies were largely conducted within the framework of the generalized I/E model (GI/EM;

Möller, Müller-Kalthoff, Helm, Nagy, & Marsh, 2016)—an extension of the I/EM to different subjects and domains. In contrast to the effects of achievements on self-concepts, it is still unknown how students' self-concept levels and changes affect their subsequent achievements, considering subject combinations other than math and native language (but see Niepel et al., 2014, Study 3). Thus, we encourage future research to further extend the R2I/EM into a generalized reciprocal 2I/E model (GR2I/EM), which would allow examining the reciprocal relations between achievement and self-concept levels and changes even more comprehensively than the R2I/EM, by considering different subject combinations.

A third limitation of our research can be seen in the self-concept scales used to test the R2I/EM. In all three studies these scales included the item “I get good grades in ...”, which has an obvious semantic overlap to students' actual grades. Thus, the inclusion of this item could have enhanced the relations between the achievement and self-concept variables, especially in Studies 1 and 2, which used grades as achievement indicators—although the self-concept item did not ask students to report their actual grades from their report cards but for their personal judgment of their subject-specific grades in general (see also Möller et al., 2011). Prior studies testing the I/EM and its extensions showed that the effects in the I/EM are not substantially affected by self-concept items referring to students' grades (e.g., Möller et al., 2011; Wolff, Nagy, et al., 2019). In line with these findings, the relations between the achievement and self-concept variables also changed very little when we repeated our analysis without the critical self-concept item (while specifying students' self-concepts as manifest variables) in Study 1 (all $|\Delta\beta| < .05$) and Study 3 (all $|\Delta\beta| < .06$).

In contrast, in Study 2 the exclusion of the self-concept item referring to students' grades led to a decrease, especially of the relations within subjects, with the strongest decreases found for the social comparison effects in the 2I/EM sub-model (math: $\Delta\beta = -.11$; German: $\Delta\beta = -.10$; all other $|\Delta\beta| \leq .07$). Accordingly, readers should note that these relations in particular may have been enhanced by the specific self-concept scale. However, there exist

only few longitudinal data sets including reciprocal assessments of achievements and self-concepts in different subjects, especially while using the same self-concept items between subjects and across time. Bearing this in mind, we considered the strong relation of one self-concept item to students' grades acceptable. Moreover, the critical item is well established in self-concept research (e.g., Marsh, 1990a, 1990b; Rost & Sparfeldt, 2002). Thus, its use allows the comparison of our findings with those of other studies. Given that our self-concept scales consisted of only three items each, we did not find it useful to exclude single items from our analyses. Nonetheless, we encourage researchers to additionally include self-concept items without reference to students' grades in future studies.

A fourth potential limitation of our research (or more generally: of most I/EM studies) concerns the fact that students reported their math and verbal self-concepts at the same measurement occasion and in close succession. This procedure could have induced context effects (e.g., Schwarz & Strack, 1991; Tourangeau & Rasinski, 1988), where students' responses to the self-concept items in one subject have influenced their responses to the self-concept items in the other. More precisely, it would be possible that the simultaneous assessment of math and verbal self-concepts has promoted the emergence of dimensional comparison effects. Helm and Möller (2017) provided at least some support for this assumption. These authors tested the classic I/EM in two groups of students who reported their math and verbal self-concepts either at the same measurement occasion or at an interval of two weeks. Although the dimensional comparison effects on students' math self-concept did not differ between the two groups, significant dimensional comparison effects on students' verbal self-concept were only found in the group of students rating both self-concepts on the same day.

To date, Helm and Möller's (2017) findings have not been replicated. Moreover, findings of a recent meta-analysis of the GI/EM speak against an enhancement of dimensional comparison effects by confronting students with self-concept items referring to multiple

subjects. In this meta-analysis, Möller, Zitzmann, Helm, Machts, and Wolff (2020) found no difference in the strength of the dimensional comparison effects in the classic I/EM between studies that asked students for their self-concepts either in math and native language only or in more than these two subjects. Nevertheless, we cannot exclude that the assessment of math and verbal self-concepts at the same measurement occasion has enhanced the dimensional comparison effects in our research. In order to counteract this possibility in the future it seems advisable to ask students for their self-concepts and other self-evaluations in different subjects with a time lag. This seems realizable in particular for data collections in large-scale studies that take place over several days. In addition, the order in which self-concepts are measured should vary randomly, which should be easy to be implemented, especially in computer-based data collections. To further investigate whether and to what extent the assessment of self-concepts in different subjects promotes the emergence of dimensional comparison effects, it would be conceivable for future studies that subgroups of students are asked to answer self-concept items in one subject only. Subsequently, it could be tested whether the relations between subject-specific achievements and self-concepts differ between these groups and the group of students responding to self-concept items in multiple subjects. To this end, researchers could make use of procedures such as FIML to deal with missing values as these would be fully explained by the study design.

Finally, one could see it as a limitation of the R2I/EM that this model does not clearly distinguish between within-person and between-person effects. Similar to its precursors (e.g., the RI/EM), the R2I/EM is a cross-lagged panel model (CLPM). Yet, several studies have shown that the autoregressive and cross-lagged effects estimated in CLPMs can be confounded with unmeasured individual and situational effects that are stable over time. As a consequence, these effects are a blend of within- and between-person variation and yield biased estimates of causal effects (e.g., Bailey, Oh, Farkas, Morgan, & Hillemeier, 2020; Berry & Willoughby, 2017; Hamaker, Kuiper, & Grasman, 2015). A potential way to address

this problem could be to incorporate random effects capturing the stable components in the achievement and self-concept constructs, similar to the random intercept cross-lagged panel model (RI-CLPM). However, in the present research, we were not able to use this approach since it requires the inclusion of each variable from at least three measurement points into one model (e.g., Hamaker et al., 2015; Mund & Nestler, 2019). Although the data analyzed in our research included achievement and self-concept variables from three measurement points, we were only able to include the data from two measurement points in the sub-models of the R2I/EM because of their temporal order. Nevertheless, we encourage researchers to consider the use of RI-CLPMs in future studies testing the R2I/EM, which allow the inclusion of achievement and/or self-concept data from more than two measurement points into one sub-model.

Conclusion

In recent years, self-concept research has yielded two longitudinal extensions of the I/EM: the RI/EM and the 2I/EM, which have specific strengths and limitations. In the present research, we developed a third longitudinal I/EM extension, the R2I/EM, to combine the strengths of both models: consideration of the reciprocal effects of self-concepts on achievements (RI/EM), and consideration of temporal comparison effects (2I/EM). Our findings, especially the differences in the effects of self-concept levels versus self-concept changes on achievements, illustrate the importance of considering temporal comparisons when analyzing the effects of self-concepts on achievements. Therefore, we hope that future research will make use of our approach, despite its relatively high complexity, to analyze the reciprocal relations between achievements and self-concepts.

References

- Albert, S. (1977). Temporal comparison theory. *Psychological Review*, 84, 485–503.
<http://dx.doi.org/10.1037/0033-295X.84.6.485>
- Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety: Within- and cross-domain relations with achievement and the mediating role of academic self-concept. *Contemporary Educational Psychology*, 51, 240–252. <http://dx.doi.org/10.1016/j.cedpsych.2017.08.005>
- Arens, A. K., Helm, F., Wolff, F., & Möller, J. (2020). Social and dimensional comparisons in the formation of German students' language self-concepts. *Journal of Experimental Education*, 88, 245–264. <http://dx.doi.org/10.1080/00220973.2019.1635069>
- Arens, A. K., & Watermann, R. (2015). How an early transition to high-ability secondary schools affects students' academic self-concept: Contrast effects, assimilation effects, and differential stability. *Learning and Individual Differences*, 37, 64–71.
<http://dx.doi.org/10.1016/j.lindif.2014.11.007>
- Bailey, D. H., Oh, Y., Farkas, G., Morgan, P., & Hillemeier, M. (2020). Reciprocal effects of reading and mathematics? Beyond the cross-lagged panel model. *Developmental Psychology*, 56, 912–921. <http://dx.doi.org/10.1037/dev0000902>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, 88, 1186–1206. <http://dx.doi.org/10.1111/cdev.12660>
- Calsyn, R., & Kenny, D. (1977). Self-concept of ability and perceived evaluations of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, 69, 136–145. <http://dx.doi.org/10.1037/0022-0663.69.2.136>

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance.

Structural Equation Modeling, 14, 464–504.

<http://dx.doi.org/10.1080/10705510701301834>

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.

<http://dx.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*, 9, 233–255.

http://dx.doi.org/10.1207/S15328007SEM0902_5

Craven, R. G., Marsh, H. W., & Burnett, P. (2004). Breaking the self-concept enhancement

conundrum: Re-conceptualising the next generation of self-concept enhancement

research. In *NZARE/AARE Conference 2003: Educational Research, Risks, and*

Dilemmas, 29 November – 3 December 2003. Auckland, New Zealand: Conference

Handbook. Retrieved from <https://eprints.qut.edu.au/26642/1/26642.pdf>

Dickhäuser, O., & Plenter, I. (2005). „Letztes Halbjahr stand ich zwei“. Zur Akkuratheit

selbst berichteter Noten. [On the accuracy of self-reported school marks.] *Zeitschrift für Pädagogische Psychologie*, 19, 219–224.

Ehm, J.-H., Hasselhorn, M., & Schmiedek, F. (2019). Analyzing the developmental relation

of academic self-concept and achievement in elementary school children: Alternative models point to different results. *Developmental Psychology*, 55, 2336–2351.

<http://dx.doi.org/10.1037/dev0000796>

Ehm, J.-H., Lindberg, S., & Hasselborn, M. (2014). Reading, writing, and math self-concept

in elementary school children: Influence of dimensional comparison processes.

European Journal of Psychology of Education, 29, 277–294.

<http://dx.doi.org/10.1007/s10212-013-0198-x>

- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
<http://dx.doi.org/10.1177/001872675400700202>
- 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. <http://dx.doi.org/10.1037/0022-0663.95.1.124>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102–116.
<http://dx.doi.org/10.1037/a0038889>
- Helm, F., & Möller, J. (2017). Der Einfluss von Erhebungsbedingungen auf Kontrasteffekte dimensionaler Vergleiche. [The influence of manner of assessment on contrastive effects of dimensional comparisons.] *Psychologie in Erziehung und Unterricht*, 64, 24–34. <http://dx.doi.org/10.1016/10.2378/peu2017.art03d>
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
<http://dx.doi.org/10.4324/9781410604118>
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- 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, 1086–1101.
<http://dx.doi.org/10.1037/edu0000021>
- Köller, O., Klemmert, H., Möller, J., & Baumert, J. (1999). Eine längsschnittliche Überprüfung des Modells des Internal/External Frame of Reference. [A longitudinal test of the internal/external frame of reference model.] *Zeitschrift für Pädagogische Psychologie*, 13, 128–134. <http://dx.doi.org/10.1024//1010-0652.13.3.128>

- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1, 259–269. <http://dx.doi.org/10.1177/2515245918770963>
- 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. <http://dx.doi.org/10.1177/0165025407077757>
- Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, 13, 59–72. http://dx.doi.org/10.1207/s15328007sem1301_3
- Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology*, 30, 263–285. <http://dx.doi.org/10.1016/j.cedpsych.2004.10.002>
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129–149. <http://dx.doi.org/10.3102/00028312023001129>
- Marsh, H. W. (1990a). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychological Review*, 2, 77–172. <http://dx.doi.org/10.1007/BF01322177>
- Marsh, H. W. (1990b). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology*, 82, 107–116. <http://dx.doi.org/10.1037/0022-0663.82.1.107>
- Marsh, H. W. (1992a). *Self-Description Questionnaire (SDQ) I: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept. A test manual and research monograph*. Macarthur: University of Western Sydney, Faculty of Education.

- Marsh, H. W. (1992b). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept. An interim test manual and research monograph*. Macarthur: University of Western Sydney, Faculty of Education.
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163.
<http://dx.doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56–67. <http://dx.doi.org/10.1037/0022-0663.96.1.56>
- Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of east and west German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology, 29*, 264–282.
[http://dx.doi.org/10.1016/S0361-476X\(03\)00034-1](http://dx.doi.org/10.1016/S0361-476X(03)00034-1)
- Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology, 39*, 326–341.
<http://dx.doi.org/10.1016/j.cedpsych.2014.08.003>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59–77.
<http://dx.doi.org/10.1348/000709910X503501>
- Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of

- lurking jingle-jangle fallacies. *Journal of Educational Psychology*, 111, 331–353.
<http://dx.doi.org/10.1037/edu0000281>
- Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin*, 34, 542–552. <http://dx.doi.org/10.1177/0146167207312313>
- Marsh, H. W., Seaton, M., Dicke, T., Parker, P. D., & Horwood, M. S. (2019). The centrality of academic self-concept to motivation and learning. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge handbook of motivation and learning* (pp. 36–62). Cambridge, UK: Cambridge University Press.
- Marsh, H. W., & Yeung, A. S. (1997a). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, 89, 41–54. <http://dx.doi.org/10.1037/0022-0663.89.1.41>
- Marsh, H. W., & Yeung, A. S. (1997b). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal*, 34, 691–720.
<http://dx.doi.org/10.2307/1163354>
- 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.
<http://dx.doi.org/10.3102/00028312035004705>
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
<http://dx.doi.org/10.1007/BF02294746>
- Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional comparisons and their consequences for self-concept, motivation, and emotion. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., pp. 430–436). Oxford, UK: Elsevier.

- Möller, J., & Husemann, N. (2006). Internal comparisons in everyday life. *Journal of Educational Psychology*, 98, 342–353. <http://dx.doi.org/10.1037/0022-0663.98.2.342>
- Möller, J., & Köller, O. (2001). Frame of reference effects following the announcement of exam results. *Contemporary Educational Psychology*, 26, 277–287. <http://dx.doi.org/10.1006/ceps.2000.1055>
- Möller, J., Kuska, S. K., & Zaunbauer, A. C. M. (2011). Internale und externe Bezugsrahmen in der Grundschule. [Internal and external frame of reference effects in elementary school.] In F. Hellmich (Ed.), *Selbstkonzepte im Grundschulalter: Modelle, empirische Ergebnisse, pädagogische Konsequenzen* [Self-concepts at elementary school age: Models, empirical findings, and educational implications] (pp. 161–172). Stuttgart, Germany: Kohlhammer.
- Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120, 544–560. <http://dx.doi.org/10.1037/a0032459>
- Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N., & Marsh, H. W. (2016). The generalized internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research*, 4(2), 1–11. <http://dx.doi.org/10.14786/flr.v4i2.169>
- 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. <http://dx.doi.org/10.1348/000709909X485234>
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79, 1129–1167. <http://dx.doi.org/10.3102/0034654309337522>
- 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. <http://dx.doi.org/10.3102/0002831211419649>
- Möller, J., Wolff, F., & Helm, F. (in press). Dimensional comparison theory: New models, new methods, new insights. In T. Dicke, F. Guay, H. W. Marsh, R. C. Craven, & D. M. McInerney (Eds.), *International advances in self research monograph, Vol. 7: Self – a multidisciplinary concept*. Charlotte, NC: Information Age Publishing.
- 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. <http://dx.doi.org/10.1111/bjep.12047>
- Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research*, 90, 376–419. <http://dx.doi.org/10.3102/0034654320919354>
- Müller-Kalthoff, H., Helm, F., & Möller, J. (2017). The Big Three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Social Psychology of Education*, 20, 849–873. <http://dx.doi.org/10.1007/s11218-017-9395-9>
- Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analyzing interdependencies across the life course. *Advances in Life Course Research*, 41, 100249. <http://dx.doi.org/10.1016/j.alcr.2018.10.002>
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, 12, 323–345. <http://dx.doi.org/10.1080/13803610600765687>

- 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. <http://dx.doi.org/10.1037/a0036307>
- Olson, B. D., & Evans, D. L. (1999). The role of the Big Five personality dimensions in the direction and affective consequences of everyday social comparison. *Personality and Social Psychological Bulletin, 25*, 1498–1508. <http://dx.doi.org/10.1177/01461672992510006>
- Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multi-context study. *Developmental Psychology, 48*, 1629–1642. <http://dx.doi.org/10.1037/a0029167>
- 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. <http://dx.doi.org/10.1016/j.cedpsych.2014.12.003>
- Raykov, T., & Marcoulides, G. A. (2009). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling, 11*, 621–637. http://dx.doi.org/10.1207/s15328007sem1104_7
- Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept—Testing the reciprocal effects model. *Learning and Instruction, 29*, 21–30. <http://dx.doi.org/10.1016/j.learninstruc.2013.07.004>
- Rost, D. H., & Sparfeldt, J. R. (2002). Facetten des schulischen Selbstkonzepts. Ein Verfahren zur Messung des differentiellen Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter). [Facets of academic self-concept. Development of a self-

- concept grid: psychometric properties and some validity data.] *Diagnostica*, 48, 130–140. <http://dx.doi.org/10.1026//0012-1924.48.3.130>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514.
<http://dx.doi.org/10.1007/BF02296192>
- Schneider, R., & Sparfeldt, J. R. (2016). Zur (Un-)Genauigkeit selbstberichteter Zensuren bei Grundschulkindern. [The accuracy of self-reported grades in elementary school.] *Psychologie in Erziehung und Unterricht*, 63, 48–59.
<http://dx.doi.org/10.2378/peu2016.art05d>
- Schurtz, I. M., Pfof, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, 34, 32–41.
<http://dx.doi.org/10.1016/j.learninstruc.2014.08.001>
- Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. *European Journal of Social Psychology*, 2, 31–50.
<http://dx.doi.org/10.1080/14792779143000015>
- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, 26, 1–2.
<http://dx.doi.org/10.1027/1015-5759/a000001>
- 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*, 58, 204–212.
<http://dx.doi.org/10.1016/j.cedpsych.2019.03.009>
- Tietjens, M., Möller, J., & Pohlmann, B. (2005). Zum Zusammenhang von Leistungen und Selbstkonzepten in verschiedenen Sportarten. [Correlations between performance and

- perceived abilities in different sports.] *Zeitschrift für Sportpsychologie*, 12, 135–143.
<http://dx.doi.org/10.1026/1612-5010.12.4.135>
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299–314.
<http://dx.doi.org/10.1037/0033-2909.103.3.299>
- Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 187–214). Berlin, Germany: Springer.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-belief and academic achievement: A meta-analytic review. *Educational Psychologist*, 39, 111–133. http://dx.doi.org/10.1207/s15326985ep3902_3
- Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development*, 89, e138–e156. <http://dx.doi.org/10.1111/cdev.12806>
- Weidinger, A. F., Steinmayr, R., & Spinath, B. (2019). Ability self-concept formation in elementary school: No dimensional comparison effects across time. *Developmental Psychology*, 55, 1005–1018. <http://dx.doi.org/10.1037/dev0000695>
- Weiner (1986). *An attributional theory of motivation and emotion*. New York, NY: Springer.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: Guilford Press.
- Wigfield, A., Eccles, J. S., Iver, D. M., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, 27, 552–565. <http://dx.doi.org/10.1037/0012-1649.27.4.552>

- Wolff, F., Helm, F., Junge, F., & Möller, J. (2020). Are dimensional comparisons performed unconsciously? An investigation of the internal/external frame of reference model using implicit self-concepts. *Journal of Educational Psychology, 112*, 397–415.
<http://dx.doi.org/10.1037/edu0000375>
- Wolff, F., Helm, F., & Möller, J. (2018). Testing the dimensional comparison theory: When do students prefer dimensional comparisons to social and temporal comparisons? *Social Psychology of Education, 21*, 875–895. <http://dx.doi.org/10.1007/s11218-018-9441-2>
- Wolff, F., Helm, F., & Möller, J. (2019). Integrating the 2I/E model into dimensional comparison theory: Towards a comprehensive comparison theory of academic self-concept formation. *Learning and Instruction, 62*, 64–75.
<http://dx.doi.org/10.1016/j.learninstruc.2019.05.007>
- Wolff, F., Helm, F., Preckel, F., & Möller, J. (2018). Dimensionale Vergleiche im Studium? Ein generalisiertes I/E-Modell zur Ausbildung fachbezogener Selbstkonzepte im Studium der Psychologie. [Dimensional comparisons at university? A generalized I/E model for the formation of subject-specific self-concepts in the study of psychology.] *Zeitschrift für Pädagogische Psychologie, 32*, 207–220.
<http://dx.doi.org/10.1024/1010-0652/a000227>
- 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*, 1005–1025. <http://dx.doi.org/10.1037/edu0000248>
- Wolff, F., Nagy, G., Retelsdorf, J., Helm, F., Köller, O., & Möller, J. (2019). The 2I/E model: Integrating temporal comparisons into the internal/external frame of reference model. *Journal of Educational Psychology, 111*, 1131–1161.
<http://dx.doi.org/10.1037/edu0000319>

- 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.
<http://dx.doi.org/10.1016/j.learninstruc.2017.09.006>
- Wolff, F., Wigfield, A., Möller, J., Dicke, A.-L., & Eccles, J. S. (2019). Social, dimensional, and temporal comparisons by students and parents: An investigation of the 2I/E model at the transition from elementary to junior high school. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000440>
- Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*, 167–202. <http://dx.doi.org/10.1111/0081-1750.00>

Table 1

Fit Indices for Self-Concept Measurement Models With Increasing Degree of Invariance Across Time

Study	Subject	Invariance	χ^2	df	p	$\Delta\chi^2$	Δdf	p	CFI	ΔCFI	RMSEA	$\Delta RMSEA$
1	Math	Configural	92.73	15	.00				.987		.059	
		Metric	99.49	19	.00	7.72	4	.10	.987	.000	.053	-.006
		Scalar	111.41	23	.00	12.17	4	.02	.986	-.001	.050	-.003
	German	Configural	59.76	15	.00				.992		.044	
		Metric	71.25	19	.00	10.89	4	.03	.990	-.002	.043	-.001
		Scalar	85.80	23	.00	14.57	4	.01	.988	-.002	.043	.000
	2	Math	Configural	66.11	15	.00			.984		.067	
		Metric	80.43	19	.00	13.91	4	.01	.981	-.003	.066	-.001
		Scalar	100.30	23	.00	20.81	4	.00	.976	-.005	.067	.001
2	German	Configural	46.24	15	.00				.988		.053	
		Metric	48.69	19	.00	1.98	4	.74	.989	.001	.046	-.007
		Scalar	54.45	23	.00	5.28	4	.26	.988	-.001	.043	-.003
	3	Math	Configural	79.03	15	.00			.995		.034	
		Metric	148.05	19	.00	71.32	4	.00	.990	-.005	.043	.009
		Scalar	205.22	23	.00	58.90	4	.00	.986	-.004	.046	.003
	Dutch	Configural	34.85	15	.00				.999		.019	
		Metric	56.72	19	.00	21.97	4	.00	.997	-.002	.023	.004
		Scalar	118.96	23	.00	68.14	4	.00	.993	-.004	.034	.011

Note. In the model assuming configural invariance, the indicators are allocated to time-specific factors, but their loadings are estimated freely across time. In the model assuming metric measurement invariance, the factor loadings are constrained to be equal over time. In the model assuming scalar invariance, the factor loadings and the intercepts of the indicators are constrained to be equal over time. Measurement invariance is indicated by the absence of changes in CFI below $-.01$ and in RMSEA over $.015$ (Chen, 2007; Cheung & Rensvold, 2002). $\Delta\chi^2$ values are Satorra-Bentler corrected (Satorra & Bentler, 2001) due to the use of the MLR estimator. Study 1: $N = 1,508$. Study 2: $N = 748$. Study 3: $N = 3,710$.

Table 2

Means, Standard Deviations, and Correlations in Study 1

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Math Achievement T1	4.33	0.92 ^{***}	1											
2. Math Achievement T3	4.03	0.88 ^{***}	.31 ^{***}	1										
3. Math Achievement T5	3.99	0.93 ^{***}	.33 ^{***}	.49 ^{***}	1									
4. German Achievement T1	4.28	0.85 ^{***}	.55 ^{***}	.17 ^{***}	.20 ^{***}	1								
5. German Achievement T3	3.98	0.88 ^{***}	.23 ^{***}	.46 ^{***}	.32 ^{***}	.37 ^{***}	1							
6. German Achievement T5	3.97	0.90 ^{***}	.19 ^{***}	.28 ^{***}	.50 ^{***}	.31 ^{***}	.50 ^{***}	1						
7. Math Self-Concept T2	2.95	0.78 ^{***}	.54 ^{***}	.32 ^{***}	.30 ^{***}	.12 ^{***}	.04	.05	1					
8. Math Self-Concept T4	2.77	0.81 ^{***}	.26 ^{***}	.65 ^{***}	.47 ^{***}	−.03	.13 ^{***}	.12 ^{***}	.51 ^{***}	1				
9. Math Self-Concept T6	2.71	0.79 ^{***}	.30 ^{***}	.44 ^{***}	.65 ^{***}	.03	.08 ^{**}	.17 ^{***}	.46 ^{***}	.64 ^{***}	1			
10. German Self-Concept T2	2.69	0.74 ^{***}	.16 ^{***}	.05	.04	.53 ^{***}	.34 ^{***}	.30 ^{***}	.13 ^{***}	−.06	−.04	1		
11. German Self-Concept T4	2.52	0.76 ^{***}	.08	.14 ^{***}	.12 ^{***}	.33 ^{***}	.63 ^{***}	.47 ^{***}	−.01	.06	−.02	.50 ^{***}	1	
12. German Self-Concept T6	2.51	0.76 ^{***}	.03	.09 ^{**}	.15 ^{***}	.28 ^{***}	.47 ^{***}	.62 ^{***}	−.05	−.02	.04	.43 ^{***}	.66 ^{***}	1

Note. The values are taken from the model which incorporates the achievement and self-concept variables from all six measurement points and also accounts for missing values. $\chi^2(181) = 868.62$, $p < .001$; CFI = .961; TLI = .940; RMSEA = .050. $N = 1,508$.

** $p < .01$. *** $p < .001$.

Table 3
Standardized Path Coefficients (and Standard Errors) in the R2I/EM in Study 1

2I/EM			r2I/EM		
Predictors	Math Self-Concept T6	German Self-Concept T6	Predictors	Math Achievement T5	German Achievement T5
Math Self-Concept T2	.25*** (.03)	−.04 (.03)	Math Achievement T1	.14* (.06)	.09 (.05)
German Self-Concept T2	−.04 (.03)	.21*** (.03)	German Achievement T1	.16** (.05)	.14* (.06)
Math Achievement Level (T3–T5)	.63*** (.03)	−.22*** (.03)	Math Self-Concept Level (T2–T4)	.36*** (.04)	.03 (.03)
German Achievement Level (T3–T5)	−.20*** (.03)	.66*** (.03)	German Self-Concept Level (T2–T4)	−.01 (.03)	.36*** (.04)
Math Achievement Change (T3–T5)	.21*** (.03)	−.02 (.02)	Math Self-Concept Change (T2–T4)	.20*** (.04)	.10** (.03)
German Achievement Change (T3–T5)	.02 (.03)	.16*** (.03)	German Self-Concept Change (T2–T4)	.08 (.04)	.17*** (.05)

Note. 2I/EM: $\chi^2(76) = 378.96, p < .001$; CFI = .969; TLI = .950; RMSEA = .051. r2I/EM: $\chi^2(76) = 453.86, p < .001$; CFI = .964; TLI = .944; RMSEA = .057. $N = 1,508$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Means, Standard Deviations, and Correlations in Study 2

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Math Achievement T1	4.54	0.76***	1											
2. Math Achievement T3	4.39	0.81***	.58***	1										
3. Math Achievement T5	4.37	0.83***	.61***	.68***	1									
4. German Achievement T1	4.68	0.45***	.24***	.15**	.19***	1								
5. German Achievement T3	4.66	0.51***	.22***	.29***	.28***	.49***	1							
6. German Achievement T5	4.73	0.53***	.19***	.16***	.22***	.55***	.52***	1						
7. Math Self-Concept T2	3.15	1.02***	.78***	.56***	.53***	.02	.02	-.05	1					
8. Math Self-Concept T4	2.97	1.03***	.61***	.78***	.62***	.07	.13*	.02	.75***	1				
9. Math Self-Concept T6	2.95	1.02***	.62***	.68***	.79***	.01	.10	-.00	.74***	.79***	1			
10. German Self-Concept T2	3.35	0.79***	-.10*	-.05	-.09	.63***	.41***	.36***	-.15**	-.08	-.17**	1		
11. German Self-Concept T4	3.24	0.82***	-.11*	-.07	-.09	.46***	.59***	.41***	-.15**	-.09	-.19***	.69***	1	
12. German Self-Concept T6	3.26	0.89***	-.12*	-.16**	-.14**	.50***	.45***	.62***	-.22***	-.16**	-.26***	.65***	.79***	1

Note. The values are taken from the model which incorporates the achievement and self-concept variables from all six measurement points and also accounts for missing values. $\chi^2(181) = 586.49$, $p < .001$; CFI = .956; TLI = .932; RMSEA = .055. $N = 748$.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
Standardized Path Coefficients (and Standard Errors) in the R2I/EM in Study 2

2I/EM			r2I/EM		
Predictors	Math Self-Concept T6	German Self-Concept T6	Predictors	Math Achievement T5	German Achievement T5
Math Self-Concept T2	.41*** (.06)	.04 (.05)	Math Achievement T1	.39*** (.10)	.32*** (.08)
German Self-Concept T2	.00 (.04)	.44*** (.04)	German Achievement T1	.16* (.07)	.42*** (.06)
Math Achievement Level (T3–T5)	.59*** (.04)	–.30*** (.04)	Math Self-Concept Level (T2–T4)	.29** (.08)	–.26** (.08)
German Achievement Level (T3–T5)	–.13** (.04)	.50*** (.04)	German Self-Concept Level (T2–T4)	–.14** (.05)	.15* (.06)
Math Achievement Change (T3–T5)	.16*** (.03)	.03 (.04)	Math Self-Concept Change (T2–T4)	.22*** (.05)	.13** (.05)
German Achievement Change (T3–T5)	–.02 (.03)	.16*** (.04)	German Self-Concept Change (T2–T4)	.06 (.04)	.18** (.06)

Note. 2I/EM: $\chi^2(76) = 262.99, p < .001$; CFI = .965; TLI = .945; RMSEA = .057. r2I/EM: $\chi^2(76) = 311.47, p < .001$; CFI = .958; TLI = .934; RMSEA = .064. $N = 748$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6
Means, Standard Deviations, and Correlations in Study 3

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Math Achievement T1	97.69	7.95***	1											
2. Math Achievement T2	103.17	8.35***	.86***	1										
3. Math Achievement T3	106.81	8.47***	.81***	.86***	1									
4. Dutch Achievement T1	49.75	7.36***	.59***	.55***	.55***	1								
5. Dutch Achievement T2	53.84	7.12***	.59***	.60***	.60***	.80***	1							
6. Dutch Achievement T3	59.01	7.84***	.60***	.60***	.64***	.77***	.83***	1						
7. Math Self-Concept T1	3.76	0.90***	.54***	.52***	.52***	.22***	.23***	.25***	1					
8. Math Self-Concept T2	3.61	0.98***	.56***	.57***	.58***	.25***	.28***	.30***	.74***	1				
9. Math Self-Concept T3	3.62	0.99***	.58***	.61***	.64***	.28***	.31***	.35***	.65***	.77***	1			
10. Dutch Self-Concept T1	3.69	0.86***	.07**	.06**	.07**	.25***	.25***	.25***	.10***	.03	.00	1		
11. Dutch Self-Concept T2	3.53	0.91***	.03	.05	.06*	.25***	.27***	.25***	.01	.07**	.00	.61***	1	
12. Dutch Self-Concept T3	3.66	0.83***	.10***	.10***	.13***	.30***	.33***	.33***	.06**	.07**	.11***	.53***	.62***	1

Note. The values are taken from the model which incorporates the achievement and self-concept variables from all three measurement points and also accounts for missing values. $\chi^2(181) = 761.24, p < .001$; CFI = .987; TLI = .981; RMSEA = .029. $N = 3,710$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 7
Standardized Path Coefficients (and Standard Errors) in the R2I/EM in Study 3

2I/EM			r2I/EM		
Predictors	Math Self-Concept T3	Dutch Self-Concept T3	Predictors	Math Achievement T3	Dutch Achievement T3
Math Self-Concept T1	.42*** (.02)	−.01 (.03)	Math Achievement T1	.61*** (.02)	.25*** (.02)
Dutch Self-Concept T1	−.06*** (.02)	.46*** (.02)	Dutch Achievement T1	.14*** (.02)	.60*** (.02)
Math Achievement Level (T2–T3)	.45*** (.03)	−.11** (.03)	Math Self-Concept Level (T1–T2)	.18*** (.02)	−.01 (.02)
Dutch Achievement Level (T2–T3)	−.05** (.02)	.29*** (.03)	Dutch Self-Concept Level (T1–T2)	−.02 (.01)	.10*** (.02)
Math Achievement Change (T2–T3)	.06*** (.02)	.04* (.02)	Math Self-Concept Change (T1–T2)	.08*** (.02)	.04* (.02)
Dutch Achievement Change (T2–T3)	.03 (.02)	−.00 (.02)	Dutch Self-Concept Change (T1–T2)	−.01 (.01)	.01 (.02)

Note. 2I/EM: $\chi^2(76) = 371.37, p < .001$; CFI = .989; TLI = .982; RMSEA = .032. r2I/EM: $\chi^2(76) = 259.32, p < .001$; CFI = .992; TLI = .987; RMSEA = .025. $N = 3,710$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 8

Summary of the Results of Hypothesis Testing in the Three Studies

Hypothesis	Study 1	Study 2	Study 3
H1: Positive effects of self-concepts on subsequent self-concepts within subjects	√	√	√
H2: Near-zero effects of self-concepts on subsequent self-concepts between subjects	√	(√)	√
H3: Positive effects of achievement levels on self-concepts within subjects	√	√	√
H4: Negative effects of achievement levels on self-concepts between subjects	√	√	√
H5: Positive effects of achievement changes on self-concepts within subjects	√	√	(√)
H6: Near-zero effects of achievement changes on self-concepts between subjects	√	√	√
H7: Positive effects of achievements on subsequent achievements within subjects	√	√	√
H8: Positive effects of achievements on subsequent achievements between subjects	(√)	√	√
H9: Positive effects of self-concept levels on achievements within subjects	√	√	√
H10: Negative effects of self-concept levels on achievements between subjects	/	√	/
H11: Positive effects of self-concept changes on achievements within subjects	√	√	(√)
H12: Near-zero effects of self-concept changes on achievements between subjects	/	/	√

Note. √ = Hypothesis supported. (√) = Hypothesis partly supported. / = Hypothesis not supported. H1–H6 refer to the 2I/EM sub-model, H7–H12 refer to the r2I/EM sub-model. Further details can be found in the running text.

Figure 1. The reciprocal effects model (REM; Marsh & Craven, 2006). The continuous lines depict effects predicted to be positive. Path labels are adapted from Niepel, Brunner, and Preckel (2014): α = skill-development effects; β = self-enhancement effect; ϵ = autoregressive effects.

Figure 2. The classic internal/external frame of reference model (I/EM; Marsh, 1986). The continuous lines depict effects predicted to be positive. The dashed lines depict effects predicted to be negative. Path labels are adapted from Niepel, Brunner, and Preckel (2014): α = social comparison effects; γ = dimensional comparison effects.

Figure 3. The reciprocal internal/external frame of reference model (RI/EM; Möller, Retelsdorf, Köller, & Marsh, 2011). The continuous lines depict effects predicted to be positive. The dashed lines depict effects predicted to be negative. The dotted lines depict effects predicted to be close to zero. Path labels are adapted from Niepel, Brunner, and Preckel (2014): α = social comparison effects / skill-development effects; β = self-enhancement effect; γ = dimensional achievement comparison effects; δ = dimensional self-concept comparison effects; ϵ = autoregressive effects; ζ = cross-domain effects between achievements; η = cross-domain effects between self-concepts.

Figure 4. The two internal/external frame of reference model (2I/EM; Wolff, Nagy, et al., 2019), controlling for prior self-concepts. The continuous lines depict effects predicted to be positive. The dashed lines depict effects predicted to be negative. The dotted lines depict effects predicted to be close to zero. Path labels are adapted from Niepel, Brunner, and Preckel (2014) and extended for the purpose of this research: α = social comparison effects / skill-development effects; γ = dimensional achievement comparison effects; ϵ = autoregressive effects; η = cross-domain effects between self-concepts; θ = temporal achievement comparison effects; τ = effects of dimensional comparisons between achievement changes.

Figure 5. The reverse two internal/external frame of reference model (r2I/EM), controlling for prior achievements. The continuous lines depict effects predicted to be positive. The dashed lines depict effects predicted to be negative. The dotted lines depict effects predicted to be close to zero. Note that after the revision of the r2I/EM (see Discussion), the dashed lines depict effects predicted to be either negative or close to zero, while the dotted lines depict effects predicted to be either close to zero or positive. Path labels are adapted from Niepel, Brunner, and Preckel (2014) and extended for the purpose of this research: β = self-enhancement effect; δ = dimensional self-concept comparison effects; ϵ = autoregressive effects; ζ = cross-domain effects between achievements; κ = temporal self-concept comparison effects; λ = effects of dimensional comparisons between self-concept changes.