



An Individual Participant Data Meta-analysis of the Joint Effects of Social, Dimensional, and Temporal Comparisons on Students' Academic Self-Concepts

Fabian Wolff¹ · Jens Möller²

Accepted: 26 May 2022 / Published online: 16 June 2022
© The Author(s) 2022

Abstract

Previous research has shown that three comparison types are involved in the formation of students' academic self-concepts: social comparisons (where students compare their achievement with their classmates), dimensional comparisons (where students compare their achievement in different subjects), and temporal comparisons (where students compare their achievement across time). The 2I/E model provides a framework to describe the joint effects of these comparisons. To date, it has been tested in 12 empirical studies. However, integration of these findings is lacking, especially in terms of yielding reliable estimates of the strength of social, dimensional, and temporal comparison effects. We therefore conducted an individual participant data (IPD) meta-analysis, in which we reanalyzed the data used in all prior 2I/E model studies ($N=45,248$). This IPD meta-analysis provided strong support for the 2I/E model: There were moderate social comparison effects, small to moderate dimensional comparison effects, and small temporal comparison effects on students' math and verbal self-concepts. Moreover, several moderating variables affected the strength of these effects. In particular, the social and temporal comparison effects were stronger in studies using grades instead of test scores as achievement indicators. Older students showed weaker social comparison effects but stronger dimensional comparison effects compared to younger students. Social comparison effects were also stronger in academic track schools compared to nonacademic track schools. Gender and migration background had only very small impacts on the strength of single comparison effects. In sum, this IPD meta-analysis significantly enhances our knowledge of comparison making in the process of students' self-concept formation.

Keywords 2I/E model · Academic self-concept · Dimensional comparisons · Social comparisons · Temporal comparisons

✉ Fabian Wolff
wolff@uni-koblenz.de

Extended author information available on the last page of the article

Subject-specific academic self-concepts can be defined as students' perceptions of their abilities in different school subjects. They are significant predictors of important educational outcomes, such as performance (e.g., Marsh & Craven, 2006), motivation (e.g., Guay et al., 2010), test anxiety (e.g., Arens et al., 2017), and course choice (e.g., Jansen et al., 2021). As such, they number among the most important constructs in educational psychology.

Among the multiple factors involved in students' academic self-concept formation, three types of comparison have been shown to be particularly relevant self-concept predictors (e.g., Wigfield et al., 2020): social comparisons (where students compare their achievement in a subject with their classmates' achievements in the same subject), dimensional comparisons (where students compare their achievement in one subject with their achievement in another subject), and temporal comparisons (where students compare their achievement in a subject with their prior achievement in that subject). Whereas social, dimensional, and temporal comparisons with lower achievements of others, in other subjects, or from former times usually increase students' self-concepts, social, dimensional, and temporal comparisons with higher achievements of others, in other subjects, or from former times usually result in self-concept decreases (e.g., Wolff et al., 2018b).

Social, dimensional, and temporal comparisons have in common that they are all embedded in a specific comparison theory: social comparison theory (Festinger, 1954), dimensional comparison theory (Möller & Marsh, 2013), or temporal comparison theory (Albert, 1977), respectively. Although the role of each of these three comparisons in the process of self-concept development has been studied for decades (e.g., Skaalvik & Skaalvik, 2004), however, it is only in recent years that researchers have focused on the simultaneous effects of social, dimensional, and temporal comparisons. In particular, this line of research was advanced by the development of the 2I/E model (Wolff et al., 2019a, b), which integrates all three comparisons and allows the examination of their joint effects on students' academic self-concepts. To date, the 2I/E model has been tested in 12 empirical studies (Wolff et al., 2018b, 2019a, b, 2020, 2021a). However, these studies differed in terms of sample characteristics (e.g., age), achievement operationalization (grades vs. test scores), and temporal comparison periods (ranging from 0.5 to 2 years). Moreover, the studies considered different numbers of subjects, and some studies included additional covariates. These differences make it difficult to compare the results over the individual studies in order to draw general conclusions about the strength of social, dimensional, and temporal comparison effects.

In the present research, we therefore bring together the 12 studies that have tested the 2I/E model, to yield reliable estimates of the strength of social, dimensional, and temporal comparison effects on students' self-concepts. Furthermore, we examine several factors that may moderate these effects. For this purpose, we conducted an individual participant data (IPD) meta-analysis in which we reanalyzed the data used in all prior 2I/E model studies. This approach—considered “the gold standard of meta-analyses” (Kaufmann et al., 2016, p. 157)—has at least three central advantages compared to classic aggregated person data (APD) meta-analyses, which synthesize data from

studies using the summary statistics presented in the respective publications (e.g., Cooper & Patall, 2009). First, the IPD approach allowed us to recalculate the 2I/E models from prior studies, while including the same variables and conducting the same statistical analyses. Thus, the results could not be biased by methodological differences between the original 2I/E model studies. At the same time, we were able to test the 2I/E model on the basis of a significantly larger sample, and thus with substantially higher power, compared to the original studies. Second, the IPD approach allowed us to examine not only moderating variables with different manifestations between studies (e.g., achievement operationalization), but also student characteristics that usually show no strong variation between studies. For example, with the IPD approach, we were able to investigate whether and how the strength of comparison effects differs between girls and boys. In contrast, in a classical APD analysis, one could only examine whether and how the strength of comparison effects depends on the gender ratio in single 2I/E model studies (which hardly differs from study to study). Finally, the IPD approach has the advantage that it avoids aggregation bias. For example, ecological fallacy, which arises when relations between variables at the group level differ from the relations between the same variables at the individual level (Robinson, 1950), can lead to misleading conclusions in APD meta-analyses. Such problems do not arise in IPD meta-analysis because it is the raw data that are analyzed.

To sum up, our meta-analysis goes beyond prior meta-analyses examining comparison effects on students' self-concepts in that it is the first that considers the joint effects of social, dimensional, and temporal comparisons, but also due to its methodological advantages. In particular, the IPD approach allows us to examine the moderating influences of variables that could not be considered in prior APD meta-analyses, such as Möller et al.'s (2020) recent meta-analysis of the joint effects of social and dimensional comparisons on students' academic self-concepts (see below). Accordingly, our research significantly enhances our knowledge not only of the strength of social, dimensional, and temporal comparison effects, but also of factors that may affect the strength of these effects in the process of students' self-concept formation.

The 2I/E Model: an Extension of the Classic I/E Model

As noted in the Introduction, the 2I/E model describes the simultaneous effects of social, dimensional, and temporal comparisons on students' math and verbal self-concepts. Wolff et al. (2019a, b) developed it on the basis of Marsh's (1986) classic internal/external frame of reference (I/E) model, which considers social comparisons (an external comparison) and dimensional comparisons (an internal comparison) in the process of self-concept formation, by adding temporal comparisons as a second internal comparison integrated into this framework. The classic I/E model and the 2I/E model share the core assumption that comparison effects can be detected by regressing students'

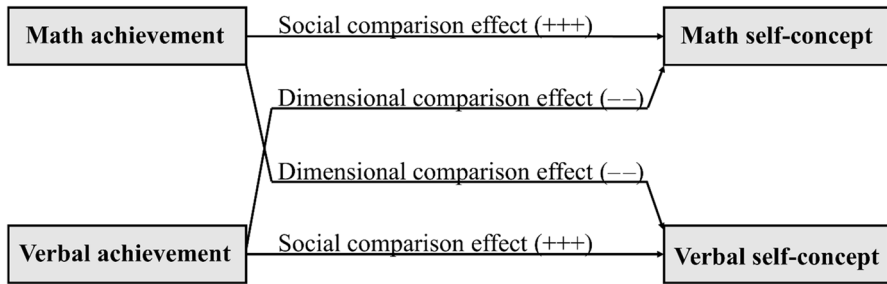


Fig. 1 The classic I/E model. The social comparison effects are assumed to be positive and stronger than the dimensional comparison effects, which are assumed to be negative. However, it should be considered that the effects of achievements on self-concepts within subjects not only result from social comparisons, but also from dimensional comparisons in particular (see section “[Additional Comments on the I/E Model and 2I/E Model](#)”)

self-concepts on their achievements. However, the approach of the 2I/E model is more complex. In the following sections, we thus briefly introduce the classic I/E model before describing the 2I/E model in detail. Subsequently, we make some comments relating to the I/E model and the 2I/E model that are of particular relevance to this meta-analysis.

The Classic I/E Model

Figure 1 shows the classic I/E model. Originally, Marsh (1986) developed it to explain why the correlation between students’ math and verbal self-concepts is usually close to zero, although students’ math and verbal achievements show strong positive correlations. Marsh argued that this seemingly paradoxical finding resulted from the joint operation of social and dimensional comparisons. On the one hand, students compared their achievement in one subject with their classmates’ achievements in the same subject. These social comparisons would lead to positive effects of achievements on self-concepts within subjects. On the other hand, students compared their math and verbal achievements with each other. In the I/E model, these dimensional comparisons would result in negative effects of achievements on self-concepts between subjects.

Empirically, researchers have examined the I/E model assumptions in more than 100 studies to date. Given this high number of studies, Möller et al. (2020) recently conducted a meta-analysis, including all studies testing the I/E model published up to April 2018 (see also Möller et al., 2009, for an earlier meta-analysis). Overall, these authors found strong support for the I/E model. Whereas the effects of students’ math achievement on their math self-concept ($\beta = .55$) and of their verbal achievement on their verbal self-concept ($\beta = .46$) were strongly positive, the effects of students’ verbal achievement on their math self-concept ($\beta = -.19$) and of their math achievement on their verbal self-concept ($\beta = -.16$) were slightly negative. In addition, experimental studies further corroborated the assumption of the I/E model that the effects of achievements on self-concepts

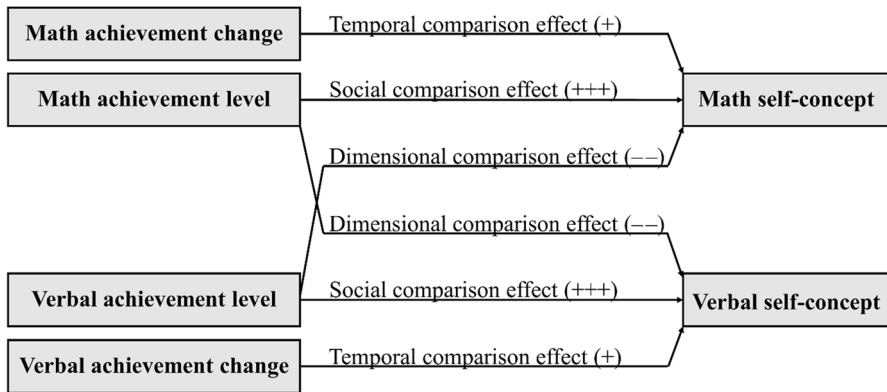


Fig. 2 The 2I/E model. The social comparison effects are assumed to be positive and stronger than the dimensional comparison effects. The dimensional comparison effects are assumed to be negative and stronger than the temporal comparison effects. The temporal comparison effects are assumed to be positive. However, it should be considered that the effects of achievement levels on self-concepts within subjects not only result from social comparisons, but also from dimensional comparisons in particular (see section “Additional Comments on the I/E Model and 2I/E Model”)

represent effects of social and dimensional comparisons (e.g., Möller & Köller, 2001; Strickhouser & Zell, 2015; Wolff & Möller, 2021).

The 2I/E Model

Although the assumptions of the classic I/E model have found empirical support in numerous studies, these studies share the limitation that they have usually neglected temporal comparisons as a third comparison type that may affect students’ self-concepts. On the basis of empirical findings from experimental, field, and vignette studies that showed an influence of temporal comparisons on students’ self-concepts beyond the influence of social and dimensional comparisons (Müller-Kalthoff et al., 2017; Wolff et al., 2018b), Wolff et al. (2019a, b) therefore extended the I/E model into the 2I/E model. Figure 2 presents this extension. As shown there, the 2I/E model describes the effects of students’ math and verbal achievements on their math and verbal self-concepts in a more differentiated fashion than the classic I/E model by considering achievement information from different points in time, and separating students’ achievement levels from their achievement changes. Students’ achievement levels are defined as their average achievements during a specific period. Students’ achievement changes refer to changes in their achievements within this period. Similarly to the classic I/E model, the 2I/E model predicts positive effects of students’ achievement levels on their self-concepts within subjects, which should result from social and dimensional comparisons, as well as negative effects of students’ achievement levels on their self-concepts between subjects, which should result from dimensional comparisons. Furthermore, the 2I/E model

predicts positive effects of students' achievement changes on their self-concepts within subjects, which should represent effects of temporal comparisons.¹

The 2I/E model makes predictions not only about the direction of the comparison effects, but also concerning their strength: According to the model, the social comparison effects should be stronger than the dimensional comparison effects, and the dimensional comparison effects should be stronger than the temporal comparison effects. On the one hand, these predictions are based on empirical findings on the joint effects of social, dimensional, and temporal comparisons from studies using other methodological approaches (e.g., Wolff et al., 2018b). On the other hand, they stem from theoretical deliberations concerning the significance of social, dimensional, and temporal comparisons in students' everyday lives.

Specifically, Wolff et al., (2019b) argued that social comparisons should have the strongest impact on students' self-concepts, because this comparison type might be the most important one in our society (see also Van Yperen & Leander, 2014). For example, if we ask how well a student performs in a school subject, we usually want to know how capable this student is in relation to his or her classmates, rather than to his or her achievement in other subjects or in the past. If we want to get a job or a college place, it is crucial that we outperform our competitors, but less so that we show better performance in the relevant domain compared to another domain or to our prior performance. In line with this, teachers often assign grades according to students' performance in relation to their classmates, but usually not in relation to their grades in other subjects or to their prior grades. However, even if social comparisons were irrelevant to academic or vocational success (e.g., because the way teachers assigned their grades was exclusively criterion-based), social comparison information would still be the most salient comparison information in many contexts, and thus might affect students' self-concepts more strongly than would dimensional or temporal comparison information. In particular, in the classroom, students are continuously confronted with the performance of others. It is likely that this exposure prompts social comparisons and—in the words of Festinger (1954, p. 117)—activates students' “drive to evaluate [their] abilities” in comparison to others.

Following social comparisons, dimensional comparisons between math and verbal achievements should have the strongest effects on students' self-concepts, according to the 2I/E model. One reason for this might be the fact that dimensional comparison information should be more salient in the classroom than temporal comparison information. Every day, students receive subject-specific achievement feedback in different subjects. Moreover, their subject-specific achievements are transparently juxtaposed on their report cards. In contrast, developments in students' achievements are still addressed relatively rarely in everyday school life (Heckhausen, 1991; Rheinberg & Krug, 2005).

¹ Some 2I/E model studies also included effects of achievement changes on self-concepts between subjects, which were interpreted as resulting from dimensional comparisons between achievement changes. However, Wolff et al. (2019a) recommended dispensing with these effects, which were usually close to zero, to reduce the complexity of the 2I/E model.

Another reason for the stronger effects of dimensional comparisons in relation to temporal comparisons may be the fact that dimensional comparisons are of particular importance for students with regard to decisions about educational and vocational specialization (e.g., Nagy et al., 2006; von Keyserlingk et al., 2021). Dimensional comparisons provide students with valuable information about their strengths and weaknesses in different subjects. Consequently, they might be of high relevance when students have to make decisions in which areas to invest their limited resources (Möller & Marsh, 2013; Wolff et al., 2018a). Nevertheless, temporal comparisons may also provide important information regarding specialization decisions, as they can give students clues that they may have overestimated or underestimated their abilities in the past, thus alerting them to possible risks of failure or chances of success in the future. With this in mind, it is plausible that students also consider their achievement changes when evaluating their abilities. Still, it is likely that achievement changes do not lead to drastic changes in students' self-concepts, considering Albert's (1977) postulate of a human drive to maintain a consistent self-perception over time. This drive implies a desire for rather stable self-concepts, which also corresponds to empirical findings illustrating the stability of academic self-concepts (e.g., Wigfield et al., 1997).

As outlined in the introduction, the 2I/E model has been tested in 12 empirical studies that considered students' achievements and self-concepts in the math and verbal domain (Wolff et al., 2018b, 2019a, b, 2020, 2021a). Overall, these studies provided strong support for the model's assumptions. However, the comparison effects showed significant differences between the studies. Specifically, the effects of achievement levels on self-concepts within subjects ranged from $\beta = .29$ to $\beta = .94$, the effects of achievement levels on self-concepts between subjects ranged from $\beta = -.46$ to $\beta = -.01$, and the effects of achievement changes on self-concepts within subjects ranged from $\beta = .04$ to $\beta = .29$. In part, these differences can be explained by the inclusion of additional covariates in some studies. For example, some studies controlled for prior self-concepts in the 2I/E model, which might have increased the temporal comparison effects but decreased the social and dimensional comparison effects (cf. Wolff et al., 2020). Still, there are also other study-specific characteristics (especially: achievement operationalization and temporal comparison periods) and sample-specific characteristics (e.g., age, track) that may have influenced the strength of the comparison effects. Hence, there is a need for a comprehensive meta-analysis that integrates the existing studies of the 2I/E model to yield reliable estimates for the strength of the comparison effects and provide explanations for the differences in the comparison effects across studies. Given the differences in the model specifications between the existing 2I/E model studies, this meta-analysis should reanalyze the data of these studies, using a unified statistical approach.

Additional Comments on the I/E Model and 2I/E Model

In this section, we make a series of important comments that apply to both the I/E model and the 2I/E model and that should be kept in mind in considering this meta-analysis.

First, we emphasize again that the explanation of the emergence of the effects in the I/E model and the 2I/E model as resulting from comparison processes is an interpretation. In fact, while these models examine the relations between achievement and self-concept variables, students are not directly asked how they compare their achievements. Nevertheless, the interpretation of comparison effects in the I/E model and 2I/E model is based on extensive theoretical considerations on the representation of comparison effects in regression models by achievement variables. Moreover, this interpretation is supported by empirical findings from studies that, using other methodological approaches, have also found evidence for the joint effects of social and dimensional or social, dimensional, and temporal comparisons in the process of students' self-concept formation. Specifically, these were experimental studies in which students received manipulated achievement feedback (e.g., Möller & Köller, 2001; Wolff et al., 2018b) and field studies in which students were directly asked to compare their achievements (e.g., Müller-Kalthoff et al., 2017; Wolff et al., 2018b). In this article, therefore, we follow this interpretation of the 2I/E model that the effects of students' achievement levels and achievement changes on their self-concepts result from comparisons, although it is acknowledged that other factors may also contribute to the emergence of these effects.

Second, we would like to note that the term “effect”, used to describe the paths in the I/E model and 2I/E model, has potential to be misleading. Although there is evidence for causal influences of social, dimensional, and temporal comparisons on self-concepts from experimental studies (Müller-Kalthoff et al., 2017; Wolff et al., 2018b), studies testing the I/E model and 2I/E model—even with longitudinal data—do not allow for such causal conclusions. However, since Marsh (1986), the term effect has been used to describe the relations between achievements and self-concepts in the I/E model, and it runs through the literature on the I/E model and 2I/E model. Moreover, in the statistical sense, the relations between achievements and self-concepts can be called effects, as they represent directed relations. With this in mind, we stay with the established effect label in the present meta-analysis, while making the point that the regression analytic approach of the 2I/E model does not allow us to draw causal conclusions.

Third, we point out that neither the I/E model nor the 2I/E model claim to represent all determinants involved in the formation of students' academic self-concepts. Rather, as is usually the case with scientific models, the I/E model and 2I/E model are simplified representations of reality. Their goal is to describe the simultaneous operation of multiple comparisons in the formation of subject-specific self-concepts. Nevertheless, there are a number of other factors that play a role in students' self-concept formation, such as gender stereotypes (e.g., Nosek et al., 2002), significant others (e.g., Fredricks & Eccles, 2002), or attribution strategies (e.g., Weiner, 1986). In more complex models, such as the expectancy-value model (e.g., Wigfield et al., 2020), many of these factors are combined. However, unlike the I/E model and the 2I/E model, these models usually cannot be fully empirically tested in one single study, due to their high complexity.

Fourth, it is important to note that the classic I/E model and 2I/E model describe the effects of social, dimensional, and temporal comparisons only in relation to one particular comparison standard, although students may conduct

several comparisons of the same type with different comparison standards when they form their self-concepts. Specifically, dimensional comparisons in the classic I/E model and 2I/E model only refer to comparisons between math and verbal achievements. This specification is reasonable insofar as studies testing the I/E model and 2I/E model with different subjects have shown that dimensional comparisons between math and verbal achievements usually show the strongest effects on students' academic self-concepts (see Möller et al., 2020, for a meta-analysis). Nevertheless, these studies have also indicated that other dimensional comparisons (e.g., between achievements in two verbal subjects) are also involved in students' self-concept formation.

Temporal comparisons in the 2I/E model refer to comparisons between achievement measures at the time points considered in the particular model. In previous studies, the selection of these time points mainly resulted from the availability of data. Still, it should be kept in mind that—similarly to dimensional comparisons—students are likely to compare their current achievements with several prior achievements when assessing their self-concepts.

Finally, social comparisons in the I/E model and 2I/E model refer to comparisons of students' achievements with the achievements of all other students examined in the respective study. However, the specification of the social comparison effects in the I/E model and 2I/E model can be viewed somewhat critically, given that there are other models of academic self-concept formation in which social comparisons are specified in relation to more local reference groups. For example, in studies testing the big-fish-little-pond effect (BFLPE; e.g., Marsh et al., 2008), the mean achievement of a specific reference group (e.g., class or school) is included in the model as an additional self-concept predictor to examine the effects of social comparisons with that specific reference group (see also Guo et al., 2018, for a discussion of social comparison effects in I/E model vs. BFLPE studies). Moreover, it is worth noting that there are models of academic self-concept formation that have chosen a term other than social comparison effect to label the effects of achievements on self-concepts within subjects. For example, in the reciprocal effect model (Marsh & Craven, 2006), these effects are called skill-development effects (see also Calsyn & Kenny, 1977). Although researchers have used social comparisons to explain them (e.g., Möller et al., 2011), it is important to consider that these effects could also result from other cognitive processes, such as normative comparisons (“How good am I compared to a general standard?”). Nevertheless, in this article, we follow the tradition of the I/E model and interpret the effects of achievement levels on self-concepts within subjects in the 2I/E model as social comparison effects. We do this because we view social comparisons as the central mechanism for the emergence of these effects, given that normative standards are likely to be shaped by achievement distributions shown in social comparison groups.

However, for the present meta-analysis, we make an important modification to the 2I/E model to ensure adequate interpretation of the social comparison effects: Instead of using achievement level in one subject to predict self-concept in the other subject, we include the difference between students' math and verbal achievement levels as a self-concept predictor in the 2I/E model (see Fig. 3). This modification is necessary because otherwise, dimensional comparison effects (defined as the effects

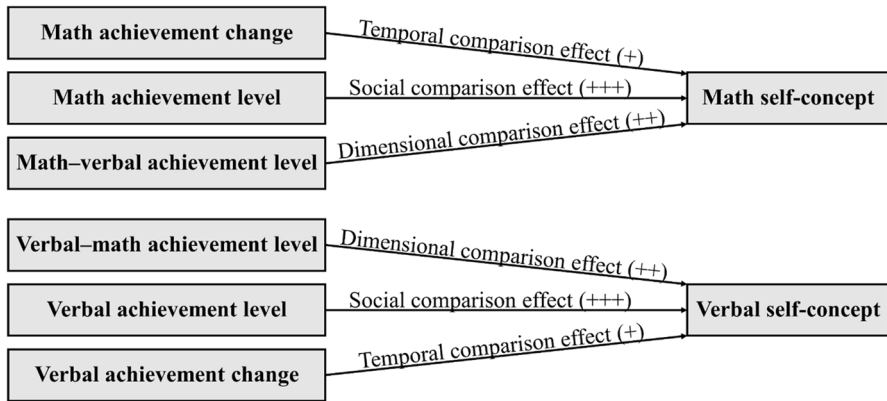


Fig. 3 The adapted 2I/E model tested in “[The Present Research](#)”. This adaptation of the 2I/E model uses the difference between students’ math and verbal achievement levels to specify the dimensional comparison effects. In contrast to the original 2I/E model, each effect of the achievement variables on the self-concept variables thus represents one specific comparison type. The social comparison effects are assumed to be stronger than the dimensional comparison effects, which are assumed to be stronger than the temporal comparison effects. Moreover, all comparison effects are assumed to be positive

of the difference between students’ math and verbal achievement levels on their self-concepts) would manifest themselves both in the effects of achievement levels on self-concepts between subjects (as intended) and in the effects of achievement levels on self-concepts within subjects (which is not intended).

Wolff (2021b) recently demonstrated this relationship for the I/E model, but his reasoning can equally be applied to the 2I/E model. Specifically, he argued that a psychologically sound operationalization of dimensional comparison effects should consider the difference between students’ verbal and math achievements as a self-concept predictor (e.g., “How much worse/better am I in math compared to English?”). In contrast, the I/E model considers students’ absolute math and verbal achievements as predictors of their self-concepts. For example, the equation $MSC = W \times MACH + B \times VACH$ (with MSC = math self-concept, $MACH$ = math achievement, $VACH$ = verbal achievement, W = I/E model effect within subjects, and B = I/E model effect between subjects) describes how in the I/E model, math self-concept is regressed on math achievement and verbal achievement. This equation is equivalent to the equation $MSC = (W + B) \times MACH + B \times (VACH - MACH)$, which includes students’ achievement difference as a self-concept predictor and thus entails adequate specification of the dimensional comparison effect. However, in this equation the social comparison effect ($W + B$) equals the sum of the (positive) effect of math achievement on math self-concept in the I/E model and the (negative) effect of verbal achievement on math achievement in the I/E model, while the dimensional comparison effect (B) is the same in both equations. Thus, in the I/E model, the effects of achievements on self-concepts within subjects represent not only effects of social comparisons but also the difference between (positive) social comparison effects and (negative) dimensional comparison effects. Concerning Möller et al.’s (2020) meta-analysis of the I/E model, for example,

this implies that the social comparison effects (i.e., the effects of achievements on self-concepts within subjects without controlling for achievement in the other subject) were only moderate: about $\beta = .36$ ($= .55-.19$) in the math domain and $\beta = .30$ ($= .46-.16$) in the verbal domain.

Similarly to the I/E model, the effects of achievement levels on self-concepts within subjects in the 2I/E model also represent not only social comparison effects (defined as effects of achievement levels on self-concepts within subjects without controlling for other variables) but the difference between (positive) social comparison effects and (negative) dimensional comparison effects. This insight, and the resulting modification of the 2I/E model (as shown in Fig. 3), is of great importance for the present meta-analysis, especially because we aim to investigate how different moderating variables influence the strength of social, dimensional, and temporal comparison effects. For this purpose, it is necessary to avoid confounding the influences of several comparison types in one model path (specifically: the influences of social and dimensional comparisons in the path from achievement levels to self-concepts within subjects), and thus allow for a clear interpretation of all comparison effects in the model. Of course, however, it is also true for the alternative specification of the 2I/E model that (as in its original specification) other factors may also contribute to the emergence of the comparison effects, as discussed above.

The Present Research

In the present research, we reanalyzed the data of all 2I/E model studies published so far in a comprehensive IPD meta-analysis. By doing so, we aimed to generate deeper knowledge about the absolute and relative strengths of social, dimensional, and temporal comparison effects in the process of students' academic self-concept formation. Moreover, we aimed to significantly enhance our knowledge of the impact of several factors that may moderate the strength of these comparison effects. In contrast to prior 2I/E model studies, we included the difference between students' math and verbal achievement levels as a self-concept predictor, so that we could isolate each effect in the 2I/E model to one specific comparative type. More specifically, we regressed students' math self-concept on their math achievement level, the difference between their math and verbal achievement levels, and their math achievement change. Similarly, we regressed students' verbal self-concept on their verbal achievement level, the difference between their verbal and math achievement levels, and their verbal achievement change (see Fig. 3). On the basis of the empirical findings of prior studies examining the joint effects of comparisons (especially those testing the 2I/E model) and the theoretical deliberations concerning the joint effects of social, dimensional, and temporal comparisons, presented above, we expected to find moderate to strong positive social comparison effects (*Hypothesis 1*), small to moderate positive dimensional comparison effects (*Hypothesis 2*), and small positive temporal comparison effects (*Hypothesis 3*) in our 2I/E model,

including data of all prior 2I/E model studies.² In addition, we made assumptions concerning the influence of particular moderating variables, which we considered in our meta-analysis.

Achievement Operationalization

In the existing 2I/E model studies, researchers have used either grades or scores from standardized tests to operationalize students' achievements. However, it is likely that the comparison effects in the 2I/E model are stronger where grades are used. Grades represent an immediate and salient form of achievement feedback for students, which can be of high importance for their future lives (e.g., concerning the transition to the next grade level or opportunities on the labor market). In contrast, students are often not even aware of their (exact) performance on standardized tests (e.g., Marsh et al., 2014b; Wolff et al., 2019b). Furthermore, teachers often assign grades as a function of the specific achievement distribution in their classes, such that the best and worst students in each class tend to get the highest and lowest grades, respectively (grading-on-a-curve). In contrast, test scores are an indicator of students' achievements in relation to a larger group of students (i.e., the population of students who took part in the achievement test). However, research on the local dominance effect (Alicke et al., 2010) has shown that individuals prefer local comparison information to more general comparison information. Accordingly, when estimating their self-concepts, students should rely on their grades more than on their results in standardized achievement tests (Marsh et al., 2014a, b; Wolff et al., 2019b).

In line with the assumption that comparison effects are stronger if grades are used as achievement indicators, Möller et al.'s (2020) meta-analysis of the classic I/E model revealed stronger I/E model effects within subjects for studies using grades rather than test scores as achievement indicators. Marsh et al. (2014b), comparing the I/E model effects depending on achievement operationalization within the same sample, even found stronger effects within and between subjects for grades as achievement indicators. Wolff et al. (2019b) were the first to compare the effects in the 2I/E model as a function of achievement operationalization, likewise in the same sample. In two studies, these authors found that most comparison effects were stronger if grades were used to operationalize students' achievements, although one-dimensional comparison effect between math and verbal subjects (an effect of student' English achievement on their math self-concept) was stronger using test scores. On the basis of these findings, we expected to find stronger social comparison effects (*Hypothesis 4*) and stronger temporal comparison effects (*Hypothesis 5*) in our meta-analysis where studies used grades rather than test scores as achievement

² We here refer to the classification of effect sizes proposed by Wolff et al. (2019b) for 2I/E model studies, according to which beta coefficients below 1.20 are classified as small effects, beta coefficients between 1.20 and 1.40 as moderate effects, and beta coefficients above 1.40 as strong effects.

indicators. However, we made no assumption concerning the impact of achievement operationalization on the strength of dimensional comparison effects, given the contradictory findings from prior research.

Comparison Period

A particularly relevant gap in the research on temporal comparisons concerns the question as to what period of time students (prefer to) consider when evaluating their achievement changes. In the 2I/E model studies, temporal comparison effects were usually operationalized depending on the available data, resulting in temporal comparison periods ranging from 0.5 to 2 years. Although the 2I/E model could be replicated while considering the different periods, however, it is conceivable that students would prefer a particular interval when comparing their achievement over time, and that different temporal comparison periods would lead to stronger or less strong temporal comparison effects in the 2I/E model. For example, it would be possible that students prefer short temporal comparison periods of one semester, because this period provides them with information about the most recent developments in their achievement. Moreover, students are more likely to remember their achievement from 6 months ago compared to that further in the past. Nevertheless, students could also prefer longer temporal comparison periods, because such periods make the occurrence of achievement changes more likely and provide students with a basis for assessing the longer-term trend of their achievement development, independent of possible short-term fluctuations. In addition, previous research has shown that social and dimensional comparison effects are stronger, the more dissimilar the individuals or subjects being compared (e.g., Huguet et al., 2009; Jansen et al., 2015; Marsh et al., 2014b; Mussweiler et al., 2004). Similarly, one could speculate that temporal comparison effects are also stronger when students compare more dissimilar (i.e., more distant) points in time. To sum up, there are arguments for students preferring either short or long comparison periods. For this reason, we left it as an open research question whether and how comparison periods affect the temporal (and social and dimensional) comparison effects in the 2I/E model.

Age

Competition in the classroom is likely to increase when students get older, and closer to applying for college and apprenticeships, where they compete with their peers. Similarly, students' need to know their strengths and weaknesses should increase with age because students usually have to specialize in one area after high school at the latest. Accordingly, it is plausible that the importance of social and dimensional comparisons would increase as students get older. Empirically, researchers have found support for this assumption (e.g., Fang et al., 2018; Möller et al., 2020). For example, in Möller et al.'s (2020) meta-analysis of the classic I/E model, the social and dimensional comparison effects

were stronger for older students compared to younger students. Furthermore, in another meta-analysis on dimensional comparisons, Wan et al. (2021) showed that the correlation between students' math and verbal self-concepts decreases with age, indicating an increase of dimensional comparison effects. In accord with these findings, we expected that students' age would also have positive effects on the strength of the social comparison effects (*Hypothesis 6*) and the dimensional comparison effects (*Hypothesis 7*) in the 2I/E model. However, we did not expect that students' age would affect the strength of the temporal comparison effects, because such comparisons should be similarly relevant for students of different age groups in keeping track of their achievement developments and considering them when assessing their self-concepts (*Hypothesis 8*).

Track

It would be reasonable for social, dimensional, and temporal comparisons to be similarly important for students in forming their self-concepts both in academic and nonacademic track schools. At both school types, students are taught together with peers in the classroom and receive achievement feedback in various subjects at regular intervals. Moreover, it is important for students in both school types to be able to assess their subject-specific abilities and thus use the comparison information available to them. In line with these deliberations, Wolff et al. (2019a, b) demonstrated the invariance of the 2I/E model effects between students of academic and nonacademic track schools in three studies. Furthermore, Arens et al., (2017, 2018), Jansen et al. (2015), and Möller et al. (2011) provided evidence for the assumption that social and dimensional comparison effects do not differ between students in academic and nonacademic track schools. In contrast to these studies, however, Möller et al. (2014) found stronger I/E model effects within subjects, indicating stronger influences of social comparisons, for students in academic track schools. Theoretically, this finding could be explained by a stronger achievement orientation in academic track schools, which would explain why students in such schools might be more concerned about their achievement and feel a greater need to evaluate their relative standing (Arens et al., 2018; Van Houtte, 2004). Given Möller et al.'s (2014) findings, we left it as an open research question whether the social comparison effects in the 2I/E model would differ between students of academic and nonacademic track schools. Nevertheless, we expected similarly strong dimensional comparison effects (*Hypothesis 9*) and temporal comparison effects (*Hypothesis 10*) for both groups of students.

Gender

Theoretically, there is little reason to believe that comparisons have different impacts on the self-concepts of girls and of boys. For both genders, it should be similarly important to know how well they perform in comparison to their peers, where their strengths and weaknesses lie, and how their performance has

developed over time (e.g., for planning their future careers). The fact that there is not much empirical evidence for gender differences in comparison effects on academic self-concepts is consistent with this reasoning. For example, Möller et al. (2020) found no difference in the I/E model effects as a function of the gender ratio in the studies included in their meta-analysis. Furthermore, several I/E model studies comparing girls and boys within the same sample demonstrated the invariance of the I/E model across gender (e.g., Arens et al., 2017; Marsh & Yeung, 1998; Möller et al., 2011; Tay et al., 1995). However, some I/E model studies have found gender differences in the comparison effects, although the pattern of findings is not consistent. For example, Skaalvik and Rankin (1990) found a stronger positive effect of verbal achievement on verbal self-concept for boys and a stronger negative effect of verbal achievement on math self-concept for girls. In contrast, Möller et al. (2014) found a stronger positive effect of math achievement on math self-concept for boys, and Nagy et al. (2006), who tested the I/E model including math and biology, found a negative effect of math achievement on biology self-concept for boys only. Given these partially contradictory findings, we abstained from formulating hypotheses concerning gender differences in the strength of social, dimensional, and temporal comparison effects, but examined the influence of gender on the comparison effects in the 2I/E model in an exploratory manner. Through our comprehensive analysis of multiple studies, we aimed to shed further light on the question of whether and how the influence of comparisons on students' self-concepts differs between girls and boys.

Migration Background

To date, researchers have not paid much attention to the question of whether students with a migration background differ from those without such a background, in the extent to which their self-concepts are influenced by comparisons. Possibly, this is due to the fact that finding differences in the strength of comparison effects between these groups does not seem very likely. In one study, Möller et al. (2014) found that the I/E model effects did not differ between students with and those without a migration background. On the basis of this finding, we also expected no influence of migration background on the strength of the social comparison effects (*Hypothesis 11*), dimensional comparison effects (*Hypothesis 12*), and temporal comparison effects (*Hypothesis 13*) in the 2I/E model.

Method

Data

In our meta-analysis, we considered all studies that have investigated the 2I/E model, including students' achievements and self-concepts in math and their first language,

and were published up to July 2021. To identify these studies, we were able to benefit from our extensive knowledge of the literature on this field of research. Moreover, we conducted a literature search as a cross-check by searching for relevant studies in the PsycArticles, PsycINFO, and PSYINDEX databases, using the search term “2I/E model” OR (“social” AND “temporal” AND “dimensional”). However, this literature search did not reveal any 2I/E model studies other than those of which we were already aware.

Overall, we identified 12 studies of the 2I/E model, published in five articles (Wolff et al., 2018b, 2019a, b, 2020, 2021a). Since two studies testing the 2I/E model used both grades and test scores (Wolff et al., 2019b, Studies 1 and 2), we formally split these into two sub-studies. However, we excluded four other studies (Wolff et al., 2018b, Studies 3a and 3b, 2019a, Study 1, 2021a, Study 1) that analyzed data already examined in another study. Moreover, we excluded a sub-study that tested the 2I/E model using parents’ ratings of their children’s competencies as criteria (Wolff et al., 2020), since the focus of our meta-analysis was on comparison effects on students’ self-concepts. From two studies that tested the 2I/E model including multiple subjects (Wolff et al., 2019b, Studies 2 and 3), we only considered the data necessary to analyze the classic 2I/E model, including math and students’ first language. To sum up, our data preparation resulted in 10 datasets to be analyzed in our IPD meta-analysis, including $N=45,248$ cases. Table 1 provides an overview of these datasets. Figure 4 illustrates the process of study selection.

We note that we included in our meta-analysis only data that had been analyzed within the framework of the 2I/E model. We did not search for additional raw data that theoretically would have allowed us to compute a 2I/E model, because our goal was to perform a meta-analysis of existing 2I/E model studies, rather than to conduct a (barely manageable) series of primary studies for subsequent meta-analysis. This approach also had the advantage that not only the selection of studies included in our meta-analysis but also the operationalization of variables in each study (e.g., selection of items to operationalize self-concepts) was unambiguous and not subject to a degree of arbitrariness.

Variables

Self-Concept Across all studies, students’ self-concepts in math and their first language were measured using 21 different self-concept items. Given this heterogeneity, we averaged the subject-specific items within the datasets and standardized the means to obtain comparable self-concept values across the datasets. All items were coded such that higher values indicated higher self-concepts. Students’ self-concepts were usually measured at the end of the comparison period or shortly thereafter. No self-concept measurement took place later than half a year after the last achievement data were collected.

Achievement In line with the 2I/E model, we used students’ achievements in math and their first language at the beginning and at the end of the temporal comparison period to calculate their subject-specific achievement levels (by averaging

Table 1 Overview of the datasets considered in the meta-analysis

Dataset	Achievement operationalization	Comparison period	Grade level	Mean age	Academic/nonacademic track	Female/male ^a	With/without migration background ^a	Country	Sample size	Reference
1 (ELEMENT ^b)	Grades	2.0 years	6	11.1 years	0.0%/0.0%	48.1%/51.3%	32.1%/64.0%	Germany	3,169	Wolff et al. (2019b), Study 1)
2 (ELEMENT ^b)	Test scores	2.0 years	6	11.1 years	0.0%/0.0%	48.1%/51.3%	32.1%/64.0%	Germany	3,169	Wolff et al. (2019b), Study 1)
3 (KESS ^c)	Grades	1.5 years	8	14.2 years	39.5%/60.5%	49.4%/49.4%	15.4%/56.2%	Germany	14,180	Wolff et al. (2019b), Study 2)
4 (KESS ^c)	Test scores	1.5 years	8	14.2 years	39.5%/60.5%	49.4%/49.4%	15.4%/56.2%	Germany	14,180	Wolff et al. (2019b), Study 2)
5 (LATEIN ^d)	Grades	0.5 years	9–11	16.1 years	100.0%/0.0%	53.2%/46.6%	7.8%/92.1%	Germany	1,225	Wolff et al. (2019b), Study 3)
6 (LISA ^e)	Grades	1.5 years	8	13.4 years	37.7%/62.3%	49.1%/50.9%	15.1%/78.5%	Germany	1,508	Wolff et al. (2019b), Study 4)
7 (G2I/EM ^f)	Grades	1.0 years	9–11	15.7 years	91.1%/8.9%	49.7%/50.3%	10.6%/89.4%	Germany	942	Wolff et al. (2019a), Study 2)
8 (MSALT ^g)	Grades	0.5 years	7	12.5 years	0.0%/0.0%	52.6%/47.4%	8.8%/88.7%	USA	2,417	Wolff et al. (2020)
9 (SEE ^h)	Grades	1.0 years	11	17.1 years	100.0%/0.0%	56.4%/43.6%	7.2%/92.0%	Switzerland	748	Wolff et al. (2021a), Study 2)

Table 1 (continued)

Dataset	Achievement operationalization	Comparison period	Grade level	Mean age	Academic/nonacademic track	Female/male ^a	With/without migration background ^a	Country	Sample size	Reference
10 (SIBO) ^b	Test scores	1.0 years	6	11.6 years	0.0%/0.0%	49.9%/50.1%	4.0%/78.8%	Belgium	3,710	Wolff et al. (2021a), Study 3)

^aThe sum of the percentages can differ from 100 since some datasets did not provide information about the gender and migration background of all students

^bLehmann (2008). The data were made available by the Research Data Centre at the Institute for Educational Quality Improvement (FDZ at IQB)

^cBos & Gröhlich (2010)

^dArens et al. (2020)

^eRetelsdorf & Möller (2013)

^fWolff et al. (2019a)

^gEccles (1996)

^hGötz & Becker (2015)

ⁱVandenberghie et al. (2012)

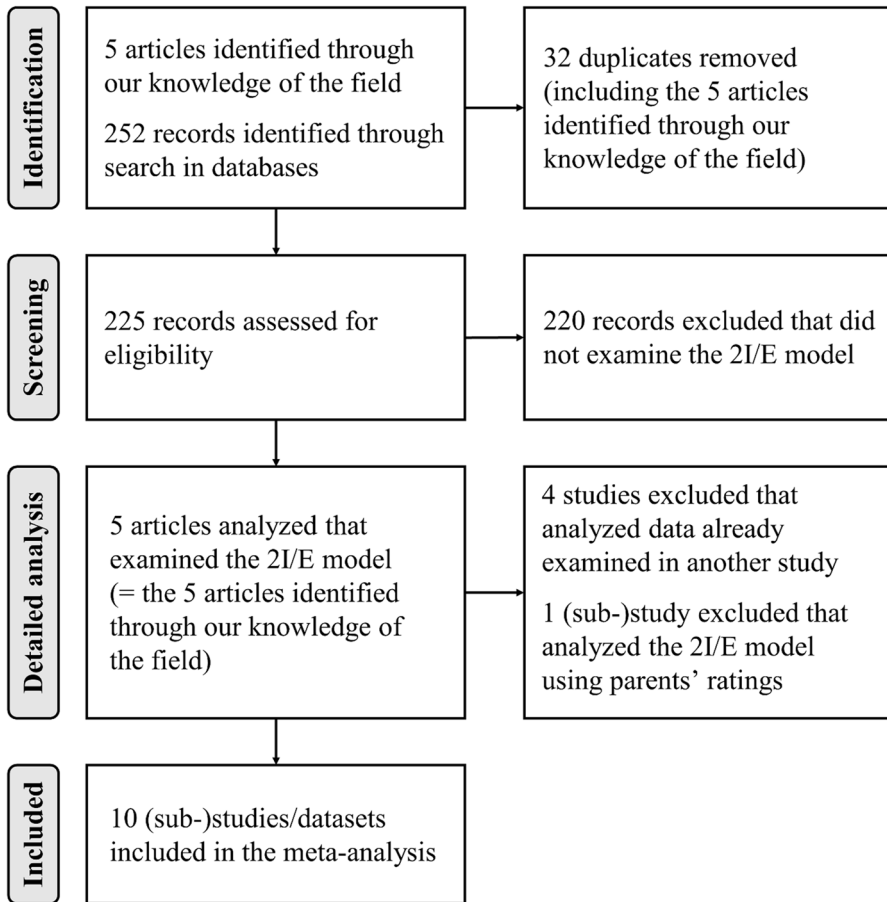


Fig. 4 The process of study selection for the meta-analysis

their subject-specific achievements) and achievement changes (by calculating the difference between their subject-specific achievements at the end and the beginning of the comparison period). Subsequently, as with students' self-concepts, we standardized their subject-specific achievement levels and achievement changes within the datasets to obtain comparable values across the datasets. All achievement scores were coded such that higher values indicated higher achievements.

Achievement Operationalization Students' achievements were measured using either grades or test scores. Thus, we coded the type of achievement operationalization as a dummy variable (0 = grades, 1 = test scores).

Comparison Period In the different studies, the temporal comparison period (i.e., the distance between the achievement measures) ranged from 0.5 to 2 years. Accordingly, we coded the length of the comparison period as a continuous variable.

Age We coded students' age at the time of self-concept assessment as a continuous variable. If age was reported in years and months, we only considered age in years, to obtain comparable data across all datasets.

Track We created three dummy variables, indicating whether the students attended a school of the academic track (0=no academic track, 1=academic track), a school of the nonacademic track (0=no nonacademic track, 1=nonacademic track), or a school in a phase of schooling or school system without tracking (0=tracking, 1=no tracking).

Gender We coded students' gender as a dummy variable (0=males, 1=females).

Migration Background We coded students' migration background as a dummy variable (0=no migration background, 1=migration background). Because migration background was measured differently in the various datasets, we had to use different criteria for coding, depending on the information available. Specifically, we coded migration background if it was noted in the dataset that the student had a migration background or was a foreigner (Datasets 1 and 2), if the student was not a citizen of the country in which the data collection took place (Datasets 3 and 4), if the student's native language was not the same as the national language of the country in which the data collection took place (Datasets 5 and 7), if at least one parent did not come from the country in which the data collection took place (Dataset 6), if the student's ethnicity was not Caucasian in a sample in which 89% of the students belonged to this ethnic group (Dataset 8), or if the student was not born in the country in which the data collection took place (Datasets 9 and 10).

Analyses

To conduct our meta-analysis, we merged the 10 single datasets. Subsequently, we used this pooled dataset to calculate six regression models in Mplus 7 (Muthén & Muthén, 2015). In the first two models, we tested the 2I/E model without moderating variables. Thus, we regressed students' math self-concept on their math achievement level, the difference between their math and verbal achievement levels, and their math achievement change (Model 1a) and students' verbal self-concept on their verbal achievement level, the difference between their verbal and math achievement levels, and their verbal achievement change (Model 1b). Afterwards, we extended these models by additionally regressing students' math self-concept (Model 2a) and verbal self-concept (Model 2b) on achievement operationalization, comparison period, age, gender, and migration background, as well as the product terms of these five moderating variables and the three achievement variables, to examine how the moderating variables affected the comparison effects. Finally, we replaced age with the two dummy variables, indicating

whether the students attended an academic track school or a nonacademic track school (i.e., no tracking served as reference category), and regressed students' math self-concept (Model 3a) and verbal self-concept (Model 3b) on the six moderating variables, the three achievement variables, and the resulting product terms of these moderating and achievement variables. To examine whether the strength of the comparison effects differed between students from academic and nonacademic track schools, we used the Model Constraint option in Mplus to compare the moderating effects of the two dummy variables, indicating an academic or a nonacademic track school. We did not include age and two tracking variables into the same model, so as to avoid problems resulting from multicollinearity because of a very strong correlation between age and students in schools without tracking (see “[Preliminary Analyses](#)” section).

To facilitate the interpretation of our results, we standardized all continuous moderating variables in the pooled dataset before calculating the product terms (students' achievement levels, achievement changes, and self-concepts had already been standardized before the single datasets were merged). Moreover, we used the standardized achievement levels to calculate the differences between the achievement levels (i.e., we subtracted two variables with the same metric from each other). For model estimation, we used the robust maximum likelihood (MLR) estimator. In addition, we used the complex modeling procedure implemented in Mplus to correct the estimated standard errors for the hierarchical structure of the data (students from different studies). To deal with missing values, we ran 20 multiple imputations in each of the single datasets before calculating the mean self-concept scores, achievement levels, and achievement changes within these datasets. Accordingly, we also conducted our analyses using 20 imputed pooled datasets. The MPlus syntax and output can be found in the Supplemental Material.

Results

Preliminary Analyses

Table 2 shows the bivariate correlations of all variables considered in our meta-analysis. In line with prior studies, the correlation between students' math and verbal self-concepts was close to zero and nonsignificant ($r = .02$, $p = .71$), whereas the correlation between their math and verbal achievement levels was strongly positive ($r = .65$, $p < .001$). Students' math and verbal achievement changes also showed a positive correlation ($r = .26$, $p < .001$).

Concerning the problem of multicollinearity in multiple regression analyses, it is important to note that most correlations among the predictor variables considered in one 2I/E model were $r \leq |.32|$. Apart from the correlations among the three dummy variables, indicating school tracking ($-.57 \leq r \leq -.43$, all $p \leq .04$), only students' achievement difference showed strong correlations

Table 2 Bivariate correlations of the variables considered in the meta-analysis

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Math self-concept	1													
2 Verbal self-concept	.02	1												
3 Math achievement level	.50***	.13***	1											
4 Verbal achievement level	.12*	.39***	.65***	1										
5 Achievement difference (math-verbal)	.45***	-.31***	.42***	-.42***	1									
6 Math achievement change	.12***	.04***	.06**	.03	.04*	1								
7 Verbal achievement change	.05**	.11***	-.00	.00	-.00	.26***	1							
8 Achievement operationalization ^a	.00	.00	.00	.00	.00	.00	.00	1						
9 Comparison period	.00	.00	.00	.00	.00	.00	.00	.14	1					
10 Age	-.02**	-.03***	-.07***	-.09***	.02	-.01***	-.00	-.16	-.24	1				
11 Academic track ^b	.04**	.07**	.29**	.32**	-.03*	.01*	-.02*	-.11	-.13	.37	1			
12 Nonacademic track ^c	-.04**	-.06**	-.28**	-.30**	.03*	-.01*	.02*	.01	.15	.34	-.57*	1		
13 No tracking ^d	.00	.00	.00	.00	.00	.00	.00	.11	-.03	-.77***	-.43**	-.51***	1	
14 Gender ^e	-.24***	.18***	-.04*	.21***	-.29***	.05***	.03	-.01	-.02*	-.02	.05***	-.04**	-.01	1
15 Migration background ^f	.01	-.04**	-.17***	-.22***	.05***	.01	.02	.01	.18*	.00	-.15***	.14*	-.00	-.01

The table shows the average results over 20 imputed datasets. All variables are standardized, except for the dummy variables and the differences between the achievement levels. The standard errors (not depicted) are corrected for the hierarchical structure of the data (students from different studies). *N* = 45,248

^a0 = grades, 1 = test scores

^b0 = no academic track, 1 = academic track

^c0 = no nonacademic track, 1 = nonacademic track

^d0 = tracking, 1 = no tracking

^e0 = males, 1 = females

^f0 = no migration background, 1 = migration background

* *p* < .05

** *p* < .01

*** *p* < .001

Table 3 Results of the meta-analysis of the 2I/E model

Predictor	Math self-concept			Verbal self-concept														
	Model 1a		Model 2a		Model 3a		Model 1b		Model 2b		Model 3b							
	B	SE	β	B	SE	β	B	SE	β	B	SE	β						
Math achievement level	.37***	.05	.37	.47***	.02	.47	.57***	.03	.57	.31***	.03	.31	.37***	.02	.37	.40***	.02	.40
Math-verbal achievement levels	.34***	.06	.29	.19***	.01	.16	.05	.04	.04	.21***	.02	.18	.20***	.01	.17	.13***	.02	.11
Math achievement change	.09***	.02	.09	.12***	.01	.12	.15***	.02	.15	.11***	.01	.11	.15***	.02	.15	.13***	.03	.13
Achievement operationalization ^a	—	—	—	-.00	.00	-.00	-.01**	.01	-.01	—	—	—	.00	.00	.00	-.01	.01	-.01
× Math achievement level	—	—	—	-.13***	.03	-.09	-.11***	.03	-.08	—	—	—	-.12***	.02	-.08	-.10***	.03	-.07
× Math-verbal achievement levels	—	—	—	.08*	.04	.04	.09*	.04	.05	—	—	—	-.04	.04	-.02	-.04	.03	-.02
× Math achievement change	—	—	—	-.07***	.01	-.05	-.07***	.01	-.05	—	—	—	-.08***	.02	-.06	-.08***	.02	-.05
Comparison period	—	—	—	-.02***	.00	-.02	-.02***	.01	-.02	—	—	—	-.01	.00	-.01	-.01*	.01	-.01
× Math achievement level	—	—	—	-.04	.03	-.04	-.02	.02	-.02	—	—	—	-.01	.01	-.01	.01	.01	.01
× Math-verbal achievement levels	—	—	—	.04	.02	.03	.02	.01	.01	—	—	—	-.01	.02	-.01	-.01	.02	-.01
× Math achievement change	—	—	—	-.01	.01	-.01	-.01	.01	-.01	—	—	—	-.00	.02	-.00	-.00	.02	-.00
Age	—	—	—	-.00	.00	-.00	—	—	—	—	—	—	-.00	.01	-.00	—	—	—
× Math achievement level	—	—	—	-.06**	.02	-.06	—	—	—	—	—	—	-.03**	.01	-.03	—	—	—
× Math-verbal achievement levels	—	—	—	.07***	.02	.06	—	—	—	—	—	—	.03*	.01	.03	—	—	—

Table 3 (continued)

Predictor	Math self-concept						Verbal self-concept								
	Model 1a			Model 2a			Model 1b			Model 2b			Model 3b		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β	B	SE	β
× Math achievement change	—	—	—	-.01*	.01	-.01	—	—	—	—	.00	.01	.00	—	—
Academic track ^b	—	—	—	—	—	—	-.08***	.01	-.03	—	—	—	—	-.07**	.03
× Math achievement level	—	—	—	—	—	—	-.09*	.04	-.06	—	—	—	—	.01	.02
× Math–verbal achievement levels	—	—	—	—	—	—	.19***	.04	.10	—	—	—	—	.09***	.03
× Math achievement change	—	—	—	—	—	—	-.05**	.02	-.03	—	—	—	—	.01	.03
Nonacademic track ^c	—	—	—	—	—	—	.04**	.01	.02	—	—	—	—	.02*	.01
× Math achievement level	—	—	—	—	—	—	-.17***	.03	-.10	—	—	—	—	-.07***	.02
× Math–verbal achievement levels	—	—	—	—	—	—	.18***	.03	.10	—	—	—	—	.07**	.02
× Math achievement change	—	—	—	—	—	—	-.03*	.01	-.02	—	—	—	—	.02	.03
Gender ^d	—	—	—	—	—	—	-.32***	.03	-.16	—	—	.12***	.01	.06	.01
× Math achievement level	—	—	—	.02*	.01	.02	.03**	.01	.02	—	—	.03	.01	.02	.02
× Math–verbal achievement levels	—	—	—	—	—	—	.07***	.02	.04	—	—	—	—	—	—
× Math achievement change	—	—	—	—	—	—	.00	.01	.01	—	—	—	—	—	—
Migration background ^e	—	—	—	—	—	—	.15***	.02	.06	—	—	.06	.04	.02	.06

Table 3 (continued)

Predictor	Math self-concept				Verbal self-concept									
	Model 1a		Model 2a		Model 1b		Model 2b							
	B	SE	B	SE	B	SE	B	SE						
× Math achievement level	—	—	-.05^{***}	.02^{***}	-.02^{***}	.01^{***}	× Verbal achievement level	—	—	-.05^{***}	.01^{***}	-.03^{***}	.02^{***}	-.02^{***}
× Math–verbal achievement levels	—	—	-.03	.01	-.01	.01	× Verbal–math achievement levels	—	—	-.00	.01	-.00	.00	.01
× Math achievement change	—	—	.01	.01	.01	.01	× Verbal achievement change	—	—	.00	.01	.00	.00	.01

The table shows the average results over 20 imputed datasets. All variables are standardized, except for the dummy variables and the differences between the achievement levels. The standard errors are corrected for the hierarchical structure of the data (students from different studies). Significant values are highlighted in bold. N = 45,248

^a0 = grades, 1 = test scores

^b0 = no academic track, 1 = academic track

^c0 = no nonacademic track, 1 = nonacademic track

^d0 = males, 1 = females

^e0 = no migration background, 1 = migration background

* $p < .05$

** $p < .01$

*** $p < .001$

with their achievement levels ($|r| = .42, p < .001$). However, these correlations were substantially weaker than the correlation between students' math and verbal achievement levels, both of which had been integrated into the same model in prior studies testing the 2I/E model. Furthermore, the correlations were reasonable, since stronger achievement differences imply more extreme absolute achievements.

As already noted in the “**Analyses**” section, there was also a very strong correlation between age and students in schools without tracking ($r = -.77, p < .001$). Since this correlation might have led to multicollinearity problems, we examined possible moderating effects of age and tracking on the comparison effects in the 2I/E model in separate analyses.

2I/E Model

Table 3 presents the results of the meta-analysis. In the first step, we considered the comparison effects in Model 1a and Model 1b, which tested the 2I/E model without moderating variables. In line with Hypothesis 1, we found moderate social comparison effects on students' math self-concept ($B = .37, p < .001, \beta = .37$) and verbal self-concept ($B = .31, p < .001, \beta = .31$). In line with Hypothesis 2, we found a moderate dimensional comparison effects on students' math self-concept ($B = .34, p < .001, \beta = .29$) and a small dimensional comparison effect on students' verbal self-concept ($B = .21, p < .001, \beta = .18$). Furthermore, we found small temporal comparison effects on students' math self-concept ($B = .09, p < .001, \beta = .09$) and verbal self-concept ($B = .11, p < .001, \beta = .11$), which supported Hypothesis 3.

In the second step, we investigated how the strength of the social, dimensional, and temporal comparison effects in the 2I/E model depended on the moderating variables considered in our meta-analysis. To this end, we examined the moderating effects for each moderating variable, except for either the tracking variables (Model 2a and Model 2b) or age (Model 3a and Model 3b). We note that the moderating effects examined in both analyses were very similar: The same moderating effects were statistically significant, and these effects hardly differed from each other (all $\Delta\beta \leq .01$). In the following, we therefore report only the results of Model 2a and Model 2b (apart from in the “**Track**” section, where we refer to the results of Model 3a and Model 3b).

Achievement Operationalization The type of achievement operationalization significantly affected the strength of all types of comparison effects. In line with Hypothesis 4, the social comparison effects were stronger for grades than for test scores (math: $B = -.13, p < .001, \beta = -.09$; verbal: $B = -.12, p < .001, \beta = -.08$). In line with Hypothesis 5, the temporal comparison effects were also stronger for grades (math: $B = -.07, p < .001, \beta = -.05$; verbal: $B = -.08, p < .001, \beta = -.06$). However, the dimensional comparison effect on students' math self-concept was stronger for test scores ($B = .08, p = .04, \beta = .04$),

whereas the strength of the dimensional comparison effect on students' verbal self-concept was not significantly affected by the type of achievement operationalization ($B = -.04, p = .30$).

Comparison Period The temporal comparison period had no significant influence on any comparison effect (all $|B| \leq .04, p \geq .09$).

Age In contrast to Hypothesis 6, we found weaker social comparison effects for older students compared to younger students (math: $B = -.06, p < .01, \beta = -.06$; verbal: $B = -.03, p < .01, \beta = -.03$). However, we found support for Hypothesis 7, since the dimensional comparison effects were stronger for older students (math: $B = .07, p < .001, \beta = .06$; verbal: $B = .03, p = .01, \beta = .03$). Moreover, Hypothesis 8 found partial support because students' age had no significant impact on the strength of the temporal comparison effect on verbal self-concept ($B = .00, p = .89$). The temporal comparison effect on math self-concept was significantly lower for older students but the effect was very small ($B = -.01, p = .02, \beta = -.01$).³

Track School track showed a number of significant influences on the strength of the comparison effects: Compared to the reference group of students in schools without tracking, we found weaker social comparison effects on math self-concept in academic track schools ($B = -.09, p = .02, \beta = -.06$) and nonacademic track schools ($B = -.17, p < .001, \beta = -.10$), as well as weaker social comparison effects on verbal self-concept in nonacademic track schools ($B = -.07, p < .001, \beta = -.04$), but not in academic track schools ($B = .01, p = .78$). The dimensional comparison effects were stronger in academic track schools (math: $B = .19, p < .001, \beta = .10$; verbal: $B = .09, p < .001, \beta = .05$) and nonacademic track schools (math: $B = .18, p < .001, \beta = .10$; verbal: $B = .07, p < .01, \beta = .04$). Moreover, there were weaker temporal comparison effects on math self-concept in academic track schools ($B = -.05, p < .01, \beta = -.03$) and nonacademic track schools ($B = -.03, p = .05, \beta = -.02$), whereas the temporal comparison effects on verbal self-concept did not differ between students in schools without tracking and schools of the academic track ($B = .01,$

³ For comparison, we also conducted our analysis including students' grade level instead of their age. Since both variables showed a very high correlation ($r = .83, p < .001$), it did not seem useful to integrate both into the same model, due to anticipated problems resulting from multicollinearity. In the additional analysis, including students' grade level, we mainly replicated our findings from Model 2a and Model 2b. All changes in the effect sizes were very small (all $|\Delta\beta| \leq .02$). However, the moderating effect of grade level on the strength of the social comparison effect on verbal self-concept was not significant ($B = -.02, p = .10$), whereas the moderating effect of age on the strength of the social comparison effect on verbal self-concept was ($B = -.03, p < .01, \beta = -.03$). Moreover, the moderating effect of gender on the strength of the social comparison effect on verbal self-concept, which hardly failed to reach statistical significance in the analysis including age ($B = .03, p = .05$), became significant in the analysis including grade level ($B = .03, p = .05, \beta = .02$).

$p = .74$) or nonacademic track ($B = .02, p = .50$). Overall, these results closely reflect the findings on the moderating effects of students' age on the strength of the comparison effects, reported in the previous section.

More interesting, therefore, was the comparison of the moderating effects between students from academic and nonacademic track schools: Here, we found stronger social comparison effects in academic track schools compared to nonacademic track schools (math: $\Delta B = .07, p < .001, \Delta\beta = .04$; verbal: $\Delta B = .08, p < .01, \Delta\beta = .05$). As predicted in Hypothesis 9, the strength of the dimensional comparison effects did not differ between academic and nonacademic track schools (math: $\Delta B = .01, p = .58$; verbal: $\Delta B = .02, p = .31$). Hypothesis 10 found partial support since the temporal comparison effects on students' verbal self-concept did not differ between academic and nonacademic track schools ($\Delta B = -.01, p = .79$), whereas the temporal comparison effects on math self-concept were—although only very slightly—weaker for students in academic track schools ($\Delta B = -.03, p < .01, \Delta\beta = -.01$). It is also interesting to note that students in nonacademic track schools had higher self-concepts compared to students in academic track schools. In fact, students in nonacademic track schools even had higher self-concepts than students in schools without tracking (math: $B = .04, p < .01, \beta = .02$; verbal: $B = .02, p = .03, \beta = .01$), whereas students in academic track schools had lower self-concepts compared to students in schools without tracking (math: $B = -.08, p < .001, \beta = -.03$; verbal: $B = -.07, p < .01, \beta = -.03$).

Gender Students' gender significantly affected the strength of the social and dimensional comparison effects on students' math self-concept: Both effects were stronger for girls compared to boys (social: $B = .02, p = .02, \beta = .02$; dimensional: $B = .07, p < .001, \beta = .04$). In contrast, students' gender had no significant impact on the strength of the other comparison effects (all $|B| \leq .03, p \geq .05$). Boys had significantly higher math self-concepts than girls ($B = -.32, p < .001, \beta = -.16$), whereas girls had significantly higher verbal self-concepts than boys ($B = .12, p < .001, \beta = .06$).

Migration Background In contrast to Hypothesis 11, the strength of the social comparison effects was significantly lower for students with a migration background compared to students without a migration background (math: $B = -.05, p < .01, \beta = -.02$; verbal: $B = -.05, p < .001, \beta = -.03$). However, in line with Hypothesis 12 and Hypothesis 13, students' migration background had no significant impact on the strength of the dimensional and temporal comparison effects (all $|B| \leq .03, p \geq .06$). Students with a migration background showed significantly higher math self-concepts compared to students without a migration background ($B = .15, p < .001, \beta = .06$).

Discussion

The present meta-analysis significantly enhances our knowledge of the joint effects of social, dimensional, and temporal comparisons in the process of academic self-concept formation. By reanalyzing the data of all existing 2I/E model studies in a comprehensive IPD meta-analysis, we were able to obtain reliable estimators of the strength of these comparison effects, on the basis of a large student sample. Moreover, we were able to gain new important insights into how various study- and student-related factors affect the strength of social, dimensional, and temporal comparison effects.

Mean Effects of Social, Dimensional, and Temporal Comparisons

In line with prior studies that had tested the 2I/E model by analyzing single datasets, our meta-analysis provided substantive empirical support for the 2I/E model. Consistent with our hypotheses, we found moderate social comparison effects, small to moderate dimensional comparison effects, and small temporal comparison effects on students' math and verbal self-concepts. However, it is remarkable that the social comparison effects, and especially the dimensional comparison effects, on students' math self-concept were significantly stronger than the social and dimensional comparison effects on their verbal self-concept (social: $\Delta\beta = .06$; dimensional: $\Delta\beta = .11$). This finding corresponds to the findings of Möller et al.'s (2020) meta-analysis of the classic I/E model, which also revealed slightly stronger social and dimensional comparison effects on students' math self-concept. It suggests that students take more account of achievement information in estimating their math self-concept than they do with their verbal self-concept. As proposed by Helm and Möller (2017), this difference in the strength of the social and dimensional comparison effects on students' math versus verbal self-concept could result from the fact that students' math self-concept is more dependent on school experiences. In their everyday lives outside of school, students usually have fewer opportunities to receive feedback on their math performance compared to their verbal performance. Therefore, they may consider achievement feedback at school more strongly when assessing their math self-concept than their verbal self-concept (see also Wolff et al., 2019b).

In contrast to the social and dimensional comparison effects, the temporal comparison effects on students' math and verbal self-concept hardly differed from each other in our meta-analysis ($\Delta\beta = -.02$). Possibly, this is because everyday life (unlike grades) typically does not provide students with nuanced feedback about their achievement changes, either in the math domain or in the verbal domain. With this in mind, it makes sense that students may consider their math and verbal achievement changes in both domains to a similar degree when assessing their self-concepts.

Moderators of Social, Dimensional, and Temporal Comparison Effects

In addition to the mean effects of social, dimensional, and temporal comparisons across the existing 2I/E model studies, our meta-analysis examined several potential moderators of social, dimensional, and temporal comparison effects. These analyses yielded a number of exciting findings. Although most of our hypotheses found empirical support, this was not the case for all of them.

Achievement Operationalization We found a high number of significant moderating effects for the type of achievement operationalization: In accord with our assumptions, the social and temporal comparison effects on students' math and verbal self-concept were stronger for studies using grades rather than test scores as achievement indicators. This finding supports the assumption that when making comparisons students consider their grades more strongly than their scores in standardized achievement tests. In particular, this may be because grades represent the more immediate and salient form of achievement feedback that students receive. The stronger social comparison effects found for grades are in accord with prior findings from studies testing the classic I/E model, especially Möller et al.'s (2020) meta-analysis. However, they go beyond these findings, since we controlled for both dimensional and temporal comparison effects in our meta-analysis. Furthermore, the finding that the temporal comparison effects were stronger in studies using grades instead of test scores as achievement indicators significantly extends research on comparison effects in the process of self-concept formation, as we have shown this for the first time in a meta-analysis.

Interestingly, we found a different picture for the impact of achievement operationalization on the strength of the dimensional comparison effects in the 2I/E model. While achievement operationalization had no significant influence on the strength of the dimensional comparison effect on students' verbal self-concept, the dimensional comparison effect on students' math self-concept was even stronger for test scores. We note that this latter finding should be interpreted with caution, since it results in particular from an exceptionally strong dimensional comparison effect on students' math self-concept in one specific 2I/E model study using test points as achievement indicators (Wolff et al., 2019b, Study 2). Nevertheless, it seems likely that dimensional comparison effects are at least not stronger for grades than they are for test scores. This conclusion is further corroborated by the results of Möller et al. (2020), who found no significant influence of achievement operationalization on the strength of the dimensional comparison effects in their meta-analysis of the classic I/E model.

It is possible that only social and temporal, but not dimensional comparison effects are stronger for grades than for test scores, because the criteria for achieving certain grades differ relatively strongly between math and verbal subjects (e.g., influence of participation in the classroom) and because math teachers often assign more extreme grades than do teachers in verbal subjects (e.g., Wolff et al., 2019b). Thus, it could be more difficult for students to compare their grades in math and verbal subjects with each other than to compare their grades in math or their first language with their classmates' grades or with their prior grades in the same subject. In

contrast, student may perceive it as relatively easy to compare their results in math and verbal achievement tests with each other if the results in both tests are presented using the same metric (e.g., percentile ranks).

Comparison Period It is interesting that we found no significant influence of the temporal comparison period on the strength of the comparison effects in the 2I/E model. Thus, it seems that students do not have a general preference for a specific period (between 0.5 and 2 years) to take recourse to when assessing their self-concepts over time. In particular, this finding is of great interest with regard to the effects of temporal comparisons, because it contradicts, among others, the assumption that there may be a distance continuum, representing the proximity of time points and relating to the strength of temporal comparison effects.

Nonetheless, readers should keep in mind that our meta-analysis compared temporal comparison periods across studies conducted with different students at different phases of their school career. For example, it is conceivable that the preference for a particular temporal comparison period would differ as a function of students' age. Students who have just transitioned from elementary to secondary school may be particularly interested in comparing their grades on their first report card at secondary school with their grades on their last report card at elementary school to see how their achievement has changed following the transition. In contrast, students at the end of high school could prefer longer comparison periods, as they tend to analyze their entire achievement development at high school. For a more specific investigation of the relation between temporal comparison periods and comparison effects, future research should, therefore, examine comparison effects as a function of different comparison periods using the same sample and compare these relations across student samples of different age groups. This approach would make it possible to draw even more substantiated conclusions regarding a potential influence of the temporal comparison period on the strength of comparison effects. Nonetheless, on the basis of the findings of our meta-analysis, we would conclude that such an influence does not exist or is at most negligibly small.

Age Our investigation of the strength of the comparison effects in the 2I/E model as a function of age revealed one unexpected result. In contrast to our assumptions, we found weaker rather than stronger social comparison effects for older students. At first glance, this finding seems counterintuitive. However, it is more reasonable when one takes into account that we only examined students from grade 6 onward. On the one hand then, our sample did not include students in their first years of schooling, for whom previous studies have found relatively weak social comparison effects (Möller et al., 2020). On the other hand, the youngest students in our sample were in the phase of transition from elementary to secondary school, and research has shown that social comparisons play a particularly important role in this phase, when students have to reevaluate their competencies in a new social comparison group (e.g., Dijkstra et al., 2008; Feldlaufer et al., 1988).

Unlike the social comparison effects, we found an increase in the dimensional comparison effects as a function of students' age. This finding was in line with our

assumptions. It underpins the assumption that dimensional comparisons become more important with greater age, when students are increasingly forced to choose between different domains—for example, when they have to decide on a profession (e.g., von Keyserlingk et al., 2021).

Also consistently with our assumptions, age showed no significant effect on the strength of the temporal effect on students' verbal self-concept. Furthermore, the strength of the temporal comparison effect on students' math self-concept hardly depended on students' age: Although this moderating effect was statistically significant (implying stronger temporal comparison effects on verbal self-concepts for younger students), it was very close to zero ($\beta = -.01$). We therefore conclude that overall, temporal comparisons have a similarly strong impact on the self-concepts of students in different age groups.

Track Our comparison of students from academic and nonacademic track schools yielded relevant differences in the comparison effects only for social comparisons. In line with Möller et al. (2014), the social comparison effects were stronger for students in academic track schools. This finding is consistent with our considerations presented above that a higher achievement orientation in academic track schools could encourage social comparisons.

The dimensional comparison effects did not differ statistically significantly between students from academic and nonacademic track schools. This finding is reasonable when it is considered that students of both school types need to know where their strengths and weaknesses lie, at the latest when it comes to vocational specializations.

Similarly to the dimensional comparison effects, the temporal comparison effect on students' verbal self-concept also did not differ statistically significantly between students from academic and nonacademic track schools. Moreover, the temporal comparison effect on students' math self-concept was stronger for students in non-academic track schools, but this difference between the school tracks was so small that it was of no practical significance ($\Delta\beta = -.01$). As with the moderating effect of students' age on the strength of the temporal comparison effects, we thus conclude that temporal comparisons have a similarly strong impact on the self-concepts of students in academic and nonacademic track schools.

Gender Our comparison of the comparison effects in the 2I/E model between girls and boys yielded slightly stronger social and dimensional comparison effects on girls' math self-concepts. To explain this finding, one could speculate that girls are more insecure than boys in assessing their math ability due to the ongoing debate about the objectively untenable gender stereotype that girls are less talented in math than are boys (e.g., Niepel et al., 2020; Wolff, 2021a). For this reason, girls could pay more attention to how they perform in math in relation to their peers and to their first language. However, it should be considered that the effects of gender on the strength of the comparison effects were very weak in our meta-analysis. Therefore, we conclude that social, dimensional, and temporal comparison effects differ very little at most between girls and boys. This conclusion is also in accord with findings

from studies demonstrating the invariance of the I/E model across gender (Arens et al., 2017; Marsh & Yeung, 1998; Möller et al., 2011; Tay et al., 1995).

Migration Background In contrast to our assumptions, we found slightly weaker social comparison effects for students with a migration background compared to students without a migration background. It is possible that this finding results from the fact that students with a migration background, compared to students without a migration background, tend to prefer to comparing their achievement with more specific social groups, including peers who are largely or exclusively of a migration background (cf. Huguet et al., 2001; Zander et al., 2014). However, although the moderating effects of migration background were consistent for both domains, we note that these effects were very small. As with the moderating effect of gender, we conclude therefore that social, dimensional, and temporal comparison effects differ very little at most between students with and without a migration background.

Strengths, Limitations, and Directions for Future Research

The present meta-analysis has many strengths. Perhaps most importantly, to the best of our knowledge, it is not only the first meta-analysis of the joint effects of social, dimensional, and temporal comparisons in the process of students' self-concept formation, but also the first meta-analysis of comparison effects using the IPD approach. This approach has enabled us to integrate and reanalyze the existing 2I/E model studies using the same analytical strategy, examine variables with manifestations that usually vary within studies rather than between studies, and avoid aggregation bias. Moreover, our meta-analysis—unlike prior studies testing the 2I/E and the I/E model (including prior meta-analyses of the I/E model)—used the difference between students' math and verbal achievement levels as a self-concept predictor to represent the effects of dimensional comparisons. Thus, we were able to clearly separate social and dimensional (and temporal) comparison effects from each other, to determine their absolute and relative strengths, and to examine their dependency on different moderating variables.

Despite the strengths of our meta-analysis, it also had some limitations. In particular, it is noteworthy that we pooled data from studies that differed from each other not only in respect to the variables controlled for in our analyses. Especially, the 2I/E model studies considered in our meta-analysis used different self-concept and achievement operationalizations. To address this problem, we standardized the self-concept and achievement variables. Still, it would have been beneficial if all students had responded to the same self-concept items and if students' achievements had been measured using the same rating systems, in order to further enhance comparability across the different studies.

Moreover, it would have been desirable if all studies had controlled for students' prior self-concepts, assessed before the first achievement measure. Had this been the case, we could have examined the effects of social, dimensional, and temporal comparisons not only on students' self-concepts measured at a particular time point, but also on changes in students' self-concepts. However, there were only

four 2I/E model studies that had controlled for prior self-concepts (Wolff et al., 2019b, Study 4, 2020, 2021a, Studies 2 and 3). Thus, controlling for prior self-concepts was not possible in our meta-analysis of all existing 2I/E model studies. We did however conduct an additional IPD meta-analysis of the four studies that had included students' prior self-concepts ($N=8,383$). This additional analysis revealed positive social comparison effects (math: $\beta=.39$; verbal: $\beta=.25$), dimensional comparison effects (math: $\beta=.12$; verbal: $\beta=.15$), and temporal comparison effects (math: $\beta=.15$; verbal: $\beta=.11$) even after controlling for prior self-concepts. The temporal comparison effect on students' verbal self-concept failed to reach statistical significance ($p=.11$). However, this temporal comparison effect was already nonsignificant before controlling for prior self-concepts, so the nonsignificant effect was not due to controlling for prior self-concepts. The exact results of the additional meta-analysis can be found in Table S1 in the Supplemental Material (see also Wolff et al., 2020 for a discussion of changes in the strength of the comparison effects after controlling for prior self-concepts). Taken together, these results provided additional support for the robustness of the 2I/E model and—in conjunction with the findings from the experimental studies showing significant effects of social, dimensional, and temporal comparisons on students' self-concepts (Müller-Kalthoff et al., 2017; Wolff et al., 2018b)—further justify the interpretation of the effects in the 2I/E model in a causal sense.

Another limitation of the present meta-analysis could be seen in the fact that all 2I/E model studies were conducted by the same first author. Although this did afford the advantage that the data needed to conduct the IPD meta-analysis were easily accessible, it would be desirable for future research on the 2I/E model that other author groups would also investigate this model, to exclude any potential influence of author group on the 2I/E model effects. However, it should be noted that most of the data analyzed in the existing 2I/E model studies were collected by other research groups (see Table 1). Thus, the results of these studies, and hence of our meta-analysis, are not likely to have been biased by expectancy effects during data collection.

As another desideratum for future research, we would like to encourage researchers to address the question as to which comparison period students prefer when making temporal comparisons to evaluate their abilities. For this purpose, as noted above, researchers could examine temporal comparison effects within the same sample for different comparison periods, supplemented perhaps by qualitative interviews. Moreover, it would be desirable for future research to examine potential influences of gender and migration background on the comparison effects in the 2I/E model, to test whether the moderating effects found for these characteristics in the present meta-analysis can be replicated in independent samples.

Concerning the generalizability of comparison effects, it would be beneficial to compare the 2I/E model relations across different countries. In the present meta-analysis, it was not reasonable to consider the country of data collection as a moderating variable, because there was only one country (Germany) in which data for more than one study were collected (see Table 1). However, Marsh and Hau (2004), in a comprehensive cross-cultural investigation of the classic I/E model, found substantial between-country differences in the strength of social

and dimensional comparison effects (although their main finding was that the relations of the model can be generalized across countries). These differences in the strength of the comparison effects, which could result from differences in the importance of particular comparison types between cultures and educational systems, should be investigated further, while also considering the effects of temporal comparisons (see also Marsh et al., 2015).

Finally, future research on comparison effects should consider students' intelligence, traits, beliefs, and behaviors when examining moderators of comparison effects. For example, Seaton et al. (2010) found that social comparison effects were stronger for students who were more intelligent, highly anxious, preferred working cooperatively, or used memorization as a learning strategy. Furthermore, Wolff et al. (2021b) showed that dimensional comparison effects were stronger the more dissimilar students perceived math and verbal school subjects to be and the more students believed that high math and verbal abilities were mutually exclusive. Given these findings, it seems promising to extend the investigation of student characteristics as potential moderators of comparison effects to variables other than sheer demographics. In particular, there is still a lack of research on how such variables affect the effects of temporal comparisons, which have long played a minor role in self-concept research, but—as shown in this meta-analysis—also have a significant impact on the formation of students' self-concepts.

Implications and Conclusion

The present meta-analysis provides strong evidence for the joint effects of social, dimensional, and temporal comparisons. On the basis of a very large sample, it shows that all three comparison types are involved in the formation of students' academic self-concepts. However, it also demonstrates that the strength of the comparison effects differs between the three comparison types, with decreasing effects from social comparisons to dimensional comparisons to temporal comparisons. Moreover, our meta-analysis revealed significant differences in the strength of the comparison effects as a function of the type of achievement operationalization, as well as students' age, school track, gender, and migration background. Researchers should be aware of these differences when conducting future studies examining comparison effects. In particular, they should be sensitive as to whether they are examining effects of comparisons between grades or between test scores, as significant differences, specifically in the strength of social and temporal comparison effects, are to be expected for these different types of achievement operationalization. Similarly, teachers should be aware not only of the general importance of comparisons at school but also of differences in the strength of comparison effects, particularly as a function of their students' characteristics, in order to be able to support their students' self-concepts optimally. It should be noted that the moderating effects of all moderating variables examined in our meta-analysis were weak and thus of rather low practical importance. Nevertheless, one could speculate that stable moderators, such as gender or

migration background, significantly influence students' ways of comparing their achievements, especially in the long term.

To sum up, we hope that our meta-analysis will make researchers and practitioners aware of the importance of comparisons in the process of self-concept formation, help them to understand the complexity of the joint effects of social, dimensional, and temporal comparisons more precisely, and stimulate further research in this field. Furthermore, we hope that our work will encourage educational researchers to conduct more IPD meta-analyses in the future, as these allow the identification of relations that cannot be found using traditional APD meta-analyses.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-022-09686-1>.

Acknowledgements We thank Stephen McLaren for copyediting the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Our work was funded by Grant WO 2341/1-1 from the German Research Foundation.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*, 485–503. <https://doi.org/10.1037/0033-295X.84.6.485>
- Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science*, *21*, 174–177. <https://doi.org/10.1177/0956797609357718>
- 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. <https://doi.org/10.1016/j.cedpsych.2017.08.005>
- Arens, A. K., Becker, M., & Möller, J. (2018). The internal/external frame of reference (I/E) model: Extension to five school subjects and invariance across German secondary school ability tracks. *Learning and Individual Differences*, *67*, 143–155. <https://doi.org/10.1016/j.lindif.2018.07.005>
- Arens, A. K., Helm, F., Wolff, F., & Möller, J. (2020). Social and dimensional comparisons in the formation of German students' language self-concepts. *The Journal of Experimental Education*, *88*, 245–264. <https://doi.org/10.1080/00220973.2019.1635069>
- Bos, W., & Gröhllich, C. (Eds.). (2010). *KESS 8: Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 8 [KESS 8: Students' competencies and attitudes at the end of Grade 8]*. Waxmann. <https://doi.org/10.24452/sjser.33.3.5287>

- Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*, 136–145. <https://doi.org/10.1037/0022-0663.69.2.136>
- Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, *14*, 165–176. <https://doi.org/10.1037/a0015565>
- Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparisons in the classroom: A review. *Review of Educational Research*, *78*, 828–879. <https://doi.org/10.3102/0034654308321210>
- Eccles, J. S. (1996). Study of life transitions (MSALT), 1983–1985. *Harvard Dataverse*. <https://doi.org/10.7910/DVN/VF4PIP>
- Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology*, *9*(1569), 1–11. <https://doi.org/10.3389/fpsyg.2018.01569>
- Feldlaufer, H., Midgley, C., & Eccles, J. S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high school. *Journal of Early Adolescence*, *8*, 133–156. <https://doi.org/10.1177/0272431688082003>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. <https://doi.org/10.1177/001872675400700202>
- Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*, 519–533. <https://doi.org/10.1037//0012-1649.38.4.519>
- Götz, T., & Becker, E. (2015). *Struktur und Ursachen von Lern- und Leistungsemotionen: Eine Längsschnittdanalyse zu habituellen und situativen Emotionen im Unterricht (SEE-Studie) [Structure and antecedents of learning and achievement emotions: A longitudinal analysis of habitual and situational emotions in class (SEE study)]*. Pädagogische Hochschule Thurgau. <https://www.phtg.ch/forschung/organisation/forschungsabteilung/abgeschlossene-projekte-a-z/struktur-und-ursachen-von-lern-und-leistungsemotionen-eine-laengsschnittdanalyse-zu-habituellen-und-situativen-emotionen-im-unterricht-see-studie/>. Accessed 2 June 2022.
- Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*, 644–653. <https://doi.org/10.1016/j.lindif.2010.08.001>
- Guo, J., Marsh, H. W., Parker, P. D., & Dicke, T. (2018). Cross-cultural generalizability of social and dimensional comparison effects on reading, math, and science self-concepts for primary school students using the combined PIRLS and TIMSS data. *Learning and Instruction*, *58*, 210–219. <https://doi.org/10.1016/j.learninstruc.2018.07.007>
- Heckhausen, H. (1991). *Motivation and action*. Springer.
- 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. <https://doi.org/10.1016/10.2378/peu2017.art03d>
- Huguet, P., Dumas, F., Monteil, J. M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, *31*, 557–578. <https://doi.org/10.1002/ejsp.81>
- Huguet, P., Dumas, F., Marsh, H. W., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, *97*, 156–170. <https://doi.org/10.1037/a0015558>
- 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. <https://doi.org/10.1037/edu0000021>
- Jansen, M., Becker, M., & Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *Journal of Educational Psychology*, *113*, 330–350. <https://doi.org/10.1037/edu0000524>
- Kaufmann, E., Reips, U.-D., & Merki, K. M. (2016). Avoiding methodological biases in meta-analysis. *Zeitschrift Für Psychologie*, *224*, 157–167. <https://doi.org/10.1027/2151-2604/a000251>
- Lehmann, R. (2008). *Erhebung zum Lese- und Mathematikverständnis: Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin [Survey of reading and mathematics comprehension: Development*

- from grades 4 to 6 in Berlin] (ELEMENT) (Version 1). Institute for Educational Quality Improvement (IQB). https://doi.org/10.5159/IQB_ELEMENT_v1
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129–149. <https://doi.org/10.3102/00028312023001129>
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1, 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., & Hau, K.-T. (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. <https://doi.org/10.1037/0022-0663.96.1.56>
- Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, 35, 705–738. <https://doi.org/10.3102/00028312035004705>
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, 20, 319–350. <https://doi.org/10.1007/s10648-008-9075-6>
- Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014a). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, 33, 50–66. <https://doi.org/10.1016/j.learninstruc.2014.04.002>
- Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014b). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology*, 39, 326–341. <https://doi.org/10.1016/j.cedpsych.2014.08.003>
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction*, 35, 16–32. <https://doi.org/10.1016/j.learninstruc.2014.08.005>
- Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, 93, 826–835. <https://doi.org/10.1037/0022-0663.93.4.826>
- Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120, 544–560. <https://doi.org/10.1037/a0032459>
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79, 1129–1167. <https://doi.org/10.3102/0034654309337522>
- Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48, 1315–1346. <https://doi.org/10.3102/0002831211419649>
- Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, 84, 591–611. <https://doi.org/10.1111/bjep.12047>
- Möller, J., 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. <https://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. <https://doi.org/10.1007/s11218-017-9395-9>
- Mussweiler, T., Rüter, K., & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology*, 87, 832–844. <https://doi.org/10.1037/0022-3514.87.6.832>
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). 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. <https://doi.org/10.1080/13803610600765687>
- Niepel, C., Stadler, M., & Greiff, S. (2020). Seeing is believing: Gender diversity in STEM is related to mathematics self-concept. *Journal of Educational Psychology*, 111, 1119–1130. <https://doi.org/10.1037/edu0000340>
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math \neq me. *Journal of Personality and Social Psychology*, 83, 44–59. <https://doi.org/10.1037//0022-3514.83.1.44>
- Retelsdorf, J., & Möller, J. (2013). Lesemotivation und Lesekompetenz in der Sekundarstufe I: Zentrale Ergebnisse der LISA-Studie [Reading motivation and reading competence in Secondary Education I: Central findings of the LISA study]. In C. Rosebrock & A. Bertschi-Kaufmann (Eds.), *Literalität erfassen: Bildungspolitisch, kulturell, individuell* (pp. 219–230). Beltz.
- Rheinberg, F., & Krug, S. (2005). *Motivationsförderung im Schulalltag: Psychologische Grundlagen und praktische Durchführung [Enhancing motivation in school: Psychological basics and practical implementation]*. Hogrefe.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351–357. <https://doi.org/10.2307/2087176>
- Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond effect: Generalizability and moderation—two sides of the same coin. *American Educational Research Journal*, 47, 390–433. <https://doi.org/10.3102/0002831209350493>
- Skaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. *Journal of Educational Psychology*, 82, 546–554. <https://doi.org/10.1037/0022-0663.82.3.546>
- Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles*, 50, 241–252. <https://doi.org/10.1023/B:SERS.0000015555.40976.e6>
- Strickhouser, J. E., & Zell, E. (2015). Self-evaluative effects of dimensional and social comparison. *Journal of Experimental Social Psychology*, 59, 60–66. <https://doi.org/10.1016/j.jesp.2015.03.001>
- Tay, M. P., Licht, B., & Tate, R. L. (1995). The internal/external frame of reference in adolescents' math and verbal self-concept: A generalization study. *Contemporary Educational Psychology*, 20, 392–402. <https://doi.org/10.1006/ceps.1995.1026>
- Van Houtte, M. (2004). Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education*, 110, 354–388. <https://doi.org/10.1086/422790>
- Van Yperen, N. W., & Leander, N. P. (2014). The overpowering effect of social comparison information: On the misalignment between mastery-based goals and self-evaluation criteria. *Personality and Social Psychology Bulletin*, 40, 676–688. <https://doi.org/10.1177/0146167214523475>
- Vandenberghe, N., Maes, F., & Van Damme, J. (2012). *Een introductie in de SiBO-databank [An introduction to the SiBO database]*. Steunpunt Studie- en Scholloopbanen. http://informatieportaalssl.be/SiBO-Databank/_AlgemeneIntroductie/DOC_IntroductieSiBODatabank.pdf. Accessed 2 June 2022.
- von Keyserlingk, L., Dicke, A.-L., Becker, M., & Eccles, J. S. (2021). What matters when? Social and dimensional comparisons in the context of university major choice. *AERA Open*, 7(1), 1–16. <https://doi.org/10.1177/23328584211020711>
- Wan, S., Lauermann, F., Bailey, D. H., & Eccles, J. S. (2021). When do students begin to think that one has to be either a “math person” or a “language person”? A meta-analytic review. *Psychological Bulletin*, 147, 867–889. <https://doi.org/10.1037/bul0000340>
- Weiner. (1986). *An attributional theory of motivation and emotion*. Springer.
- Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbretton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, 89, 451–469. <https://doi.org/10.1037/0022-0663.89.3.451>
- Wigfield, A., Eccles, J. S., & Möller, J. (2020). How dimensional comparisons help to understand linkages between expectancies, values, performance, and choice. *Educational Psychology Review*, 32, 657–680. <https://doi.org/10.1007/s10648-020-09524-2>

- Wolff, F. (2021a). How classmates' gender stereotypes affect students' math self-concepts: A multilevel analysis. *Frontiers in Psychology, 12*(599199), 1–12. <https://doi.org/10.3389/fpsyg.2021.599199>
- Wolff, F. (2021b). Measuring comparison effects: A critical view on the internal/external frame of reference model. *Educational Measurement: Issues and Practice, 40*, 89–95. <https://doi.org/10.1111/emip.12414>
- Wolff, F., & Möller, J. (2021). Dimensional comparison theory: Minimal intervention affects strength of dimensional comparison effects. *The Journal of Experimental Education, 89*, 625–642. <https://doi.org/10.1080/00220973.2020.1843128>
- Wolff, F., Helm, F., & Möller, J. (2018a). Testing the dimensional comparison theory: When do students prefer dimensional comparisons to social and temporal comparisons? *Social Psychology of Education, 21*, 875–895. <https://doi.org/10.1007/s11218-018-9441-2>
- Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018b). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology, 110*, 1005–1025. <https://doi.org/10.1037/edu0000248>
- Wolff, F., Helm, F., & Möller, J. (2019a). 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. <https://doi.org/10.1016/j.learninstruc.2019.05.007>
- Wolff, F., Nagy, G., Retelsdorf, J., Helm, F., Köller, O., & Möller, J. (2019b). The 2I/E model: Integrating temporal comparisons into the internal/external frame of reference model. *Journal of Educational Psychology, 111*, 1131–1161. <https://doi.org/10.1037/edu0000319>
- Wolff, F., Wigfield, A., Möller, J., Dicke, T., & Eccles, J. S. (2020). 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, 112*, 1644–1660. <https://doi.org/10.1037/edu0000440>
- Wolff, F., Sticca, F., Niepel, C., Götz, T., Van Damme, J., & Möller, J. (2021a). 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. *Journal of Educational Psychology, 113*, 1529–1549. <https://doi.org/10.1037/edu0000632>
- Wolff, F., Zitzmann, S., & Möller, J. (2021b). Moderators of dimensional comparison effects: A comprehensive replication study putting prior findings on five moderators to the test and going beyond. *Journal of Educational Psychology, 113*, 621–640. <https://doi.org/10.1037/edu0000505>
- Zander, L., Webster, G. D., & Hannover, B. (2014). Better than me?! How adolescents with and without migration background in Germany perceive each others' performance in class. *Psychology, 5*, 62–69. <https://doi.org/10.4236/psych.2014.51011>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Fabian Wolff¹ · Jens Möller²

¹ Institute of Psychology, University of Koblenz-Landau, Koblenz Campus, Universitätsstraße 1, 56070 Koblenz, Germany

² Institute for Psychology of Learning and Instruction, Kiel University, Olshausenstraße 75, 24118 Kiel, Germany