



A longitudinal experiment examining the joint effects of social, dimensional, and temporal comparisons on self- and external assessments of students' abilities

Fabian Wolff¹ 

Received: 24 October 2022 / Accepted: 7 June 2023 / Published online: 13 July 2023
© The Author(s) 2023

Abstract

This study examined the joint effects of social comparisons (comparisons of one's own achievements with those of others), dimensional comparisons (comparisons of one's own achievements in different domains), and temporal comparisons (comparisons of one's own achievements across time) on students' ability assessments. For the first time a longitudinal experiment, using $N=411$ students from German universities, was conducted to examine these comparison effects on students' own self-concepts, their inferred self-concepts of a peer, and their own assessments of a peer's abilities in the figural and verbal domains. As expected, significant social and dimensional comparison effects on the different kinds of ability ratings were found. However, contrary to assumptions, most temporal comparison effects were nonsignificant. The social comparison effects were stronger than the dimensional and temporal comparison effects. Moreover, the social comparison effects were stronger, while one dimensional comparison effect was weaker, when students assessed their peers' abilities, rather than inferring the self-concepts of their peers.

Keywords Ability assessment · Dimensional comparisons · Self-concept · Social comparisons · Temporal comparisons

1 Introduction

Self-concepts of ability are of great importance to students because they predict key outcomes such as achievement (e.g., Wolff et al., 2021), interest (e.g., Marsh et al., 2005), and career decisions (e.g., Jansen et al., 2021). Among the multiple factors involved in students' self-concept formation, recent research has emphasized the

✉ Fabian Wolff
wolff@uni-koblenz.de

¹ Institute of Psychology, University of Koblenz, Universitätsstraße 1, 56070 Koblenz, Germany

significance of three comparisons, each the subject of a specific comparison theory: social comparisons (Festinger, 1954), where students compare their achievement with those of others, dimensional comparisons (Möller & Marsh, 2013), where students compare their achievement in different domains, and temporal comparisons (Albert, 1977), where students compare their achievement across time. Social, dimensional, and temporal comparisons have in common that they relate a target (i.e., the achievement of the person, in the domain, or at the time point of interest) with a standard (i.e., the achievement of the person, in the domain, or at the time point of reference). Moreover, they share that comparisons can be directed downward (i.e., the target is superior to the standard), laterally (i.e., the target is similar to the standard), or upward (i.e., the target is inferior to the standard).

To date, several studies have suggested that social, dimensional, and temporal downward comparisons usually increase students' self-concepts, whereas social, dimensional, and temporal upward comparisons decrease students' self-concepts (e.g., Wolff & Möller, 2022). However, most of these studies are non-experimental and thus do not allow drawing causal conclusions about the impact of the comparisons. Specifically, only three experimental studies so far have examined the joint effects of social, dimensional, and temporal comparisons on students' assessments of their own abilities (Wolff et al., 2018b; Zell & Strickhouser, 2020). However, only one of these studies has found a significant effect of temporal comparisons (Wolff et al., 2018b).

One explanation for this inconsistency could be the fact that temporal comparisons in the experimental studies referred to achievement changes within a period of less than one hour, whereas temporal comparisons in the non-experimental studies referred to periods of months or even years. Therefore, in this research I conducted a longitudinal experiment in which students assessed their self-concepts after having received manipulated feedback on their achievement on two ability tests completed at two different measurement points at a distance of two weeks. Moreover, I investigated the joint effects of social, dimensional, and temporal comparisons on external ratings of ability. For this purpose, I asked students to assess a peer's ability from both the peer's perspective ("students' inferred self-concepts of a peer") and from their own perspective ("students' own assessments of a peer's ability").

Overall, this study is of great relevance in both theoretical and practical respects. From a theoretical perspective, the investigation of the joint effects of social, dimensional, and temporal comparisons is particularly important with respect to the development of a comprehensive comparison theory of academic self-concept formation. As noted above, social, dimensional, and temporal comparisons each originate in a specific comparison theory. However, given the possible joint influences of these comparisons on students' self-concepts, it would be desirable to integrate them into one comparison theory at some point. This comparison theory should also address the question of the extent to which the effects of social, dimensional, and temporal comparisons can be generalized to different kinds of ability assessment, and offer explanations for possible differences between comparison effects on self- and external assessments. To this end, it is beneficial to find out whether comparison effects on external ratings of ability depend on whether these ratings are made from the other person's perspective (i.e., are inferences

about others' self-concepts) or from one's own perspective (i.e., are one's own assessments of others' abilities), because such knowledge could explain possible differences between comparison effects on self- and external assessments of ability. For example, it would be possible that comparison effects on self- and external assessments differ because certain comparison information has different relevance in various kinds of ability ratings (e.g., as individuals also make comparisons to serve certain motivations, such as the desire to feel good; Wolff et al., 2018a). However, it is also conceivable that differences in comparison effects between self- and external assessments result from deficient abilities to put oneself in another person's shoes and to intuit this person's cognitive processes (e.g., as individuals conduct comparisons partly on an unconscious level; Wolff et al., 2020a).

In practical terms, this study is particularly relevant with regard to the role of temporal comparisons in the process of students' self-concept formation. Given that comparisons influence students' self-concepts, the question arises to what extent these comparisons could be used, for example, by teachers or parents as a means of promoting students' self-concepts. In this regard, temporal comparisons are particularly promising, as in many cases it should be relatively easy for teachers to point their students toward achievement improvements (or at least competence gains); for instance, by applying an individualized teacher frame of reference (e.g., Helm et al., 2023). In contrast, it may be more difficult for teachers to counteract undesirable influences from social and dimensional comparisons, because teachers have limited influence on achievement differences in the classroom, which are particularly predictive of social comparison effects (e.g., Marsh et al., 2014), and because dimensional comparisons are a "double-edged sword" (Möller & Marsh, 2013, p. 546) in that positive effects of dimensional comparisons on self-concept in one domain are accompanied by negative effects on self-concept in the other domain. Beyond that, studying the influence of comparisons on different kinds of ability assessments is of high practical importance. For optimal promotion of students' academic self-concepts, teachers should be able to replicate the cognitive processes their students go through when assessing their self-concepts. However, if teachers have limited or no ability to do this, it would be advisable to educate them about the influences of comparisons on different kinds of ability assessment.

2 Literature overview

2.1 Comparison effects on students' self-concepts

In recent years, the joint effects of social, dimensional, and temporal comparisons on students' self-concepts of ability have been investigated in several non-experimental studies. In particular, researchers have conducted 12 studies testing the 2I/E model (Wolff et al., 2019; see Wolff & Möller, 2022, for a meta-analysis). In this model, students' math and verbal self-concepts are regressed on their math and verbal achievement levels and achievement changes during a specific period, and the effects of achievement levels and changes on self-concepts are interpreted as social, dimensional, and temporal comparison effects, respectively. In addition, some

non-experimental studies have regressed students' math and verbal self-concepts on their direct evaluations of their math or verbal achievements in comparison to their classmates, their achievement in the other domain, and their prior achievement in the same domain (Müller-Kalthoff et al., 2017a, Study 3; Wolff et al., 2018b, Study 2). Overall, these studies suggest that social, dimensional, and temporal comparisons are involved in the formation of students' domain-specific self-concepts. Furthermore, they indicate that social comparisons have the strongest effect, followed by dimensional and then temporal comparisons. According to Wolff et al. (2018b), the relative strength of comparison effects might reflect the importance of the different types of comparison in modern societies and their salience in students' everyday lives.

However, to interpret the comparison effects in a causal sense, it is necessary to replicate the findings from non-experimental studies using experimental designs. To date, researchers have conducted several experiments examining the joint effects of either social and dimensional comparisons (Möller & Köller, 2001; Pohlmann & Möller, 2006; Strickhouser & Zell, 2015) or social and temporal comparisons (Van Yperen & Leander, 2014; Zell & Alicke, 2009, 2010) on students' self-concepts. In sum, these studies found support for the joint effects of both comparisons examined. Moreover, the social comparison effects were usually substantially stronger, which led Van Yperen and Leander (2014, p. 676) to speak about "the overpowering effect of social comparison information" (TOESCI).

Considering that the experimental studies examining two comparison types found significant effects of both comparisons on students' domain-specific self-concepts, it would be plausible that experimental studies examining all three comparison types together would similarly find significant effects of all three comparisons. However, this was shown to be the case in only one of three studies (Wolff et al., 2018b, Study 1). In this study, students rated their ability to solve figure analogy tasks after they received manipulated feedback concerning their achievement in a figure analogies test. According to this feedback, their achievement in the figure analogies test was either better or worse than the mean achievement in a reference group in this test, than their achievement in a word analogies test, and in the second part of the figure analogies test compared to the first part. Consistent with the findings from most non-experimental studies, the social comparison feedback showed the strongest effect on students' self-concept, followed by dimensional comparison feedback, then temporal comparison feedback.

Building on Wolff et al.'s (2018b) findings, Zell and Strickhouser (2020) examined the joint effects of social, dimensional, and temporal comparisons on students' self-perceived ability to solve verbal reasoning tasks in two experimental studies. In addition to a verbal reasoning test, the students also worked on a quantitative reasoning test. Study 1 was conceptually similar to Wolff et al.'s (2018b) experiment. In Study 2, the students received achievement feedback that triggered an upward or downward comparison with respect to only one comparison type, while lateral comparisons were triggered with respect to the other two comparison types. Interestingly, in both studies the only significant comparison effects were social and dimensional, while the temporal comparison effects were nonsignificant. Moreover,

in both studies the social comparison effects were significantly stronger than the other comparison effects.

One explanation for the contradictory findings between Wolff et al.'s (2018b) and Zell and Strickhouser's (2020) experiments could be that the researchers investigated comparison effects on self-concepts in different domains. However, this argument becomes less plausible considering findings from non-experimental studies revealing significant effects of all three comparison types on students' math and verbal self-concepts. Therefore, a more reasonable explanation for the lack of temporal comparison effects in Zell and Strickhouser's (2020) studies could be the fact that achievement feedback in these studies referred to achievement changes over a period of a few minutes. Although this limitation also applies to Wolff et al.'s (2018b) experiment, the explanation is strengthened by the fact that Zell and Alicke (2009, 2010) found significant social and temporal comparison effects in two studies in which students completed a bogus social sensitivity test over the course of several weeks before their social sensitivity was assessed.

In summary, Zell and Strickhouser's (2020) findings call into question the role of temporal comparisons in the process of students' self-concept formation. Accordingly, there is a need for an experimental study that examines the joint effects of social, dimensional, and temporal comparisons based on achievement feedback that addresses more than one measurement point. Ideally, such a study should examine comparison effects on self-concepts in different domains to uncover potential dependencies on the domain under investigation.

2.2 Comparison effects on external ratings of ability

The findings on the influence of comparisons on ability assessments become even more ambiguous when a distinction is made between self- and external ratings. In the only non-experimental study that has examined the effects of social, dimensional, and temporal comparisons on self- and external ratings, Wolff et al. (2020b) found significant effects of all three types of comparison on students' math and verbal self-concepts and parents' assessments of their children's math and verbal abilities. For both kinds of assessment the social comparison effects were stronger than the dimensional and temporal comparison effects. However, the effects of dimensional and temporal comparisons were significantly weaker for the external assessments.

Unlike Wolff et al. (2020b), most non-experimental studies examining the effects of social and dimensional comparisons on math and verbal ability ratings by others (Dai, 2002; Helm et al., 2018; Lösch et al., 2017; Marsh et al., 1984; Pohlmann et al., 2004; Van Zanden et al., 2017) have found no significant dimensional contrast effects (i.e., the higher ability ratings in one domain the lower the achievement in the other) or have even found dimensional assimilation effects (i.e., the higher ability ratings in one domain the higher the achievement in the other). However, it is possible that the direction and strength of dimensional comparison effects depend on the kind of ability rating. For example, Van Zanden et al. (2017) found significant dimensional contrast effects when parents were asked to infer their children's

self-concepts, whereas the dimensional comparison effects were nonsignificant when the parents rated their children's abilities from their own perspective.

In experimental studies, dimensional comparison effects on ability ratings by others have so far been investigated for inferred self-concepts only. In particular, Müller-Kalthoff et al., (2017a, Studies 1–2) examined dimensional comparison effects along with social and temporal comparison effects in two vignette studies. It is interesting that, while the social comparison effects were the strongest, the temporal comparison effects exceeded the dimensional comparison effects in both studies. This finding is consistent with the findings of Zell and Alicke (2010), who found a stronger influence of temporal upward comparisons when students assessed the social sensitivity of a peer instead of their own. However, it contradicts the findings of Zell and Alicke (2009), who found no significant temporal comparison effect when students' social sensitivity was assessed by their peers.

To sum up, the findings on the impact of comparisons on external ability assessments are rather inconsistent. In particular, there is a lack of experimental studies examining the joint effects of social, dimensional, and temporal comparisons on one's own self-concepts, one's inferred self-concepts of another person, and one's own ratings of another person's abilities.

3 The present research

This study is the first to investigate the joint effects of social, dimensional, and temporal comparisons on students' own self-concepts, their inferred self-concepts of a peer, and their own assessments of a peer's abilities in the figural and verbal domains, using a longitudinal experimental design. In particular, this approach aimed to generate substantive knowledge about the impact of temporal comparisons in students' self-concept formation and about possible differences between comparison effects on different kinds of ability assessment.

On the basis of previous findings, and given the longitudinal design of this study, I expected that social comparisons (*Hypothesis 1.1*), dimensional comparisons (*Hypothesis 2.1*), and temporal comparisons (*Hypothesis 3.1*) would show significant effects on students' own self-concepts, indicating higher self-concepts following downward and lower self-concepts following upward comparisons. I also expected that social comparisons (*Hypothesis 1.2*), dimensional comparisons (*Hypothesis 2.2*), and temporal comparisons (*Hypothesis 3.2*) would show significant effects on students' inferred self-concepts of a peer, indicating higher inferred self-concepts following downward and lower inferred self-concepts following upward comparisons. However, for students' own assessments of a peer's abilities, I only predicted significant social comparison effects (*Hypothesis 1.3*), indicating higher ability assessments following downward and lower ability assessments following upward comparisons, while examining the effects of dimensional and temporal comparisons exploratorily.

In line with TOESCI, I further expected that the social comparison effects would be significantly stronger than the dimensional comparison effects on students' own self-concepts (*Hypothesis 4.1*), on their inferred self-concepts of a peer (*Hypothesis 4.2*), and on their own assessments of a peer's abilities (*Hypothesis 4.3*). Similarly,

I assumed that the social comparison effects would be significantly stronger than the temporal comparison effects on students' own self-concepts (*Hypothesis 5.1*), on their inferred self-concepts of a peer (*Hypothesis 5.2*), and on their own assessments of a peer's abilities (*Hypothesis 5.3*).

4 Method

4.1 Overview

The study design was similar to that of Zell and Strickhouser (2020, Study 2). At two measurement points (T1 and T2), approximately two weeks apart, students worked on a figural ability test (FAT) and a verbal ability test (VAT). At T2, they received manipulated achievement feedback that triggered two lateral comparisons and one downward, lateral, or upward comparison with regard to their achievement in at least one test (see Table 1). On the basis of this feedback, the students rated their ability to solve FAT and VAT tasks (i.e., their FAT and VAT self-concepts). Similarly, the students also received feedback about the achievement of a (fictitious) peer who had supposedly worked on the two ability tests, and then rated this peer's ability to solve FAT and VAT tasks from their own perspective (inferred self-concepts of the peer) and the peer's perspective (own assessments of the peer's abilities). Data collection took place online in small groups and lasted almost 2 h per measurement point.

To examine the effects of social, dimensional, and temporal comparisons, I compared the ability ratings between the two experimental conditions triggering downward versus upward comparisons with regard to the respective comparison type, along with lateral comparisons for the other two comparison types. This approach was similar to all three previous experimental studies that investigated the joint effects of social, dimensional, and temporal comparisons, in which the researchers also examined comparison effects by comparing experimental conditions triggering upward versus downward comparisons (Wolff et al., 2018b, Study 1; Zell & Strickhouser, 2020, Studies 1 and 2). Thus, my approach allowed a direct comparison of this study's findings with those from the previous studies. In addition, I preferred the approach of operationalizing comparison effects by comparing experimental conditions triggering upward and downward comparisons to an alternative approach of comparing experimental conditions triggering upward and lateral comparisons as well as downward and lateral comparisons, because my approach was likely to have the higher power (since students' ability ratings should show stronger differences after upward versus downward comparisons than after upward versus lateral or downward versus lateral comparisons). Nevertheless, for additional analyses, the design of this study also included a condition triggering lateral comparisons for all three comparison types (Condition 1 in Table 1).

4.2 Sample

Participants were $N=433$ students from 63 German universities and universities of applied science. Of these, $n=411$ students (94.9%) participated at both

Table 1 Feedback conditions

Condition	Comparison Direction				Percentile Ranks			
	FAT		VAT		FAT		VAT	
	Social	Dimensional	Temporal	Social	Dimensional	Temporal	T1	T2
1 (“Reference”)	Lateral	Lateral	Lateral	Lateral	Lateral	Lateral	50	50
2 (“SocDown”)	Downward	Lateral	Lateral	Downward	Lateral	Lateral	70	70
3 (“SocUp”)	Upward	Lateral	Lateral	Upward	Lateral	Lateral	30	30
4 (“DimDown ^{FAT} ”)	Lateral	Downward	Lateral	Upward	Upward	Lateral	50	50
5 (“DimUp ^{FAT} ”)	Lateral	Upward	Lateral	Downward	Downward	Lateral	50	50
6 (“DimDown ^{VAT} ”)	Upward	Upward	Lateral	Lateral	Downward	Lateral	30	30
7 (“DimUp ^{VAT} ”)	Downward	Downward	Lateral	Lateral	Upward	Lateral	70	70
8 (“TemDown”)	Lateral	Lateral	Downward	Lateral	Lateral	Downward	40	60
9 (“TemUp”)	Lateral	Lateral	Upward	Lateral	Lateral	Upward	60	40

FAT = Figural Ability Test. VAT = Verbal Ability Test. T1 = First measurement point. T2 = Second measurement point. A downward/lateral/upward comparison indicates a higher/similar/lower achievement in comparison to the reference group (social comparison), to the other test (dimensional comparison), or at T2 in comparison to T1 (temporal comparison). To examine social comparison effects, Conditions 2 and 3 were compared. To examine dimensional comparison effects, Conditions 4 and 5 (FAT) or 6 and 7 (VAT) were compared. To examine temporal comparison effects, Conditions 8 and 9 were compared. All percentile ranks were varied randomly by ±2% to increase the credibility of the manipulated feedback

measurement points. The mean age in this sample was 23.3 years ($SD=4.71$). Most of the students were female (76.9%), German (96.1%), and psychology students (68.6%). Other majors that were nominated by at least five students (i.e., more than 1%) were sociology (17.5%; 15.6% in combination with psychology), management/business administration (10.5%; 7.5% in combination with psychology), German studies (6.1%; 0.2% in combination with psychology), English studies (4.9%; 1.2% in combination with psychology), education sciences/pedagogy (2.9%), mathematics (2.9%), biology (2.4%; 0.2% in combination with psychology), bio-geosciences (1.2%), computational visualistics (1.2%), philosophy (1.2%), and sport sciences (1.2%). A series of t -tests showed that psychology students and non-psychology students differed neither in their FAT or VAT achievements nor in their FAT or VAT self-concepts (all $t < 1$, all $p \geq .34$). Of the students who participated at both measurement points, between $n=114$ and $n=152$ students were further excluded from the analyses because of manipulation check failures or technical problems with data storage (see Table S1 in the Supplemental Material for detailed subsample descriptions). Still, these sample sizes allowed for the detection of small-to-medium effects of $f^2 > .04$ with a power of $> 95\%$.

Students were recruited mainly via online forum postings and announcements in lectures. As a thank for their participation, they had the chance to win vouchers of a total value of about \$3,000. Moreover, psychology students received course credits. Participation was voluntary. Participants were randomly assigned to the experimental conditions and data were collected until each condition included at least 40 participants.

4.3 Materials and variables

4.3.1 Ability tests

The ability tests were composed of the three figural (FAT) and verbal (VAT) subtests of the German Intelligence Structure Test 2000 R (I-S-T 2000 R; Liepmann et al., 2007). Each of these subtests consisted of 20 items and was available in two parallel versions for use at T1 (Version A) and T2 (Version C). Two subtests had already been used in previous experiments to examine the influence of comparisons on domain-specific self-concepts (Möller & Köller, 2001; Wolff et al., 2018b). In the present study, I included additional subtests to increase the credibility of the manipulated feedback. The correlations between students' raw scores in the different figural and verbal subtests at one measurement point ranged from $r = .33$ to $r = .51$; those between their raw scores in the same subtests at the different measurement points from $r = .45$ to $r = .67$.

4.3.2 Independent variables

The students received two sets of manipulated achievement feedback: one on their own achievement and one on a peer's achievement. Both sets of feedback were randomly assigned and independent of each other. As shown in Table 1, each set of

achievement feedback related to one of the nine experimental conditions and was given in the form of percentile ranks. These percentile ranks indicated what percentage of students in a reference group performed worse than the students themselves, or the peer, respectively. The percentile ranks were varied systematically in relation to Condition 1 (three lateral comparisons): To trigger social downward comparisons, all percentile ranks were increased by 20%, whereas all percentile ranks were decreased by 20% to trigger social upward comparisons. To trigger dimensional downward comparisons, both percentage ranks in the standard domain were decreased by 20%, whereas both percentage ranks in the standard domain were increased by 20% to trigger dimensional upward comparisons. To trigger temporal downward comparisons, both percentage ranks at T1 were decreased by 10% and both percentage ranks at T2 were increased by 10%, whereas both percentage ranks at T1 were increased by 10% and both percentage ranks at T2 were decreased by 10% to trigger temporal upward comparisons. Thus, each upward and downward comparison referred to an achievement difference of 20% (between the student's/peer's achievement and the average achievement in the reference group, between the achievement in the FAT and the achievement in the VAT, or between the achievement at T1 and the achievement at T2). This achievement difference was consistent with Zell and Strickhouser (2020, Study 2). However, to increase the credibility of the manipulated feedback, all percentile ranks were also varied randomly by up to 2%. Figure 1 shows an example of the feedback a student could have received during the study.

4.3.3 Dependent variables

Students' domain-specific own self-concepts, inferred self-concepts of a peer, and own assessments of a peer's abilities were measured using six analogous items that had been used successfully in previous self-concept research (e.g., Wolff et al., 2018b). These items were adapted slightly for the three different kinds of ability assessment. For example, one item read "Tasks like those in the figural/verbal ability test are easy for me" (students' own self-concept), "My fellow student thinks: Tasks like those in the figural/verbal ability test are easy for me" (students' inferred self-concept of a peer), or "Tasks like those in the figural/verbal ability test are easy for my fellow student" (students' own assessment of a peer's ability), respectively. Students responded to each item on a Likert scale ranging from 1 = *strongly disagree* to 6 = *strongly agree*. Reliabilities of all scales were very high (all $.90 \leq \alpha \leq .94$). A complete list of all items and reliabilities can be found in Table S2 in the Supplemental Material.

4.3.4 Manipulation checks

As a manipulation check, students were asked to rate their own and the peer's achievement in the FAT and VAT in comparison to the reference group, to the other test, and at T2 compared to T1, on a Likert scale ranging from 1 = *much worse* to 9 = *much better* (see Table S3 in the Supplemental Material for exact item formulations). Students were excluded from the respective analyses if they

Your achievement in the figural ability test	
1 st measurement point: 71%	2 nd measurement point: 70%
Your achievement in the verbal ability test	
1 st measurement point: 69%	2 nd measurement point: 72%
Note: The percentile ranks indicate how many students of the comparison group performed worse than you in the respective test. Example: A percentile rank of 60 would mean that you did better than 60% of the students in the comparison group.	
Your fellow student's achievement in the figural ability test	
1 st measurement point: 51%	2 nd measurement point: 50%
Your fellow student's achievement in the verbal ability test	
1 st measurement point: 29%	2 nd measurement point: 28%
Note: The percentile ranks indicate how many students of the comparison group performed worse than your fellow student in the respective test. Example: A percentile rank of 60 would mean that your fellow student did better than 60% of the students in the comparison group.	

Fig. 1 Example of achievement feedback. The feedback on the student's own achievement (above) and the achievement of the peer (below) was presented sequentially and in random order. The order of feedback about achievement in the FAT and in the VAT also varied. For each set of feedback, one of the nine experimental conditions was randomly selected. Moreover, each percentile rank was randomly varied by up to 2%. In the example shown, the feedback on the student's own achievement was based on Condition 2. The feedback on the achievement of the peer was based on Condition 4. This feedback has been freely translated from German into English

did not tick values above 5 when downward comparisons were induced, below 5 when upward comparisons were induced, or between 3 and 7 when lateral comparisons were induced. Overall, this led to the exclusion of $n = 152$ students (37.0%) in the analyses examining the comparison effects on students' FAT self-concept, $n = 136$ students (33.1%) in the analyses examining the comparison effects on students' VAT self-concept, $n = 113$ students (27.5%) in the analyses examining the comparison effects on students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks, and $n = 117$ students (28.5%) in the analyses examining the comparison effects on students' inferred VAT self-concept of a peer and students' own assessment of a peer's ability to solve VAT tasks.

Table 2 presents how many students responded to the single items used for the manipulation check in a way that was consistent with the experimental manipulation. As shown, students responded to the items more consistently with the manipulation if the items referred to achievement ratings in temporal comparison rather than social comparison ($\Delta = 4.0\%$) or dimensional comparison ($\Delta = 10.8\%$), and to achievement ratings in social comparison rather than dimensional comparison ($\Delta = 6.8\%$). Furthermore, the students responded to the items more consistently with the manipulation if the items referred to their peer's achievement rather than their own achievement ($\Delta = 4.4\%$). The domain to which the items referred was

Table 2 Percentage of students responding to single items used for the manipulation check consistently with the experimental manipulation

Person	Domain	Comparison			Mean
		Social	Dimensional	Temporal	
Self	Figural	83.9%	79.1%	89.3%	84.1%
	Verbal	86.4%	81.0%	90.0%	85.8%
Other	Figural	91.2%	82.5%	94.6%	89.5%
	Verbal	90.8%	82.7%	94.4%	89.3%
Mean Self		85.2%	80.0%	89.7%	85.0%
Mean Other		91.0%	82.6%	94.5%	89.4%
Mean Figural		87.6%	80.8%	92.0%	86.8%
Mean Verbal		88.6%	81.9%	92.2%	87.6%
Mean		88.1%	81.3%	92.1%	87.2%

The items used for the manipulation check asked the students to assess their own (“Self”) and their peer’s (“Other”) achievement in the Figural Ability Test (“Figural”) and in the Verbal Ability Test (“Verbal”) in comparison to the reference group (“Social”), to the other test (“Dimensional”), and at second measurement point compared to the first measurement point (“Temporal”), on a Likert scale ranging from 1 = *much worse* to 9 = *much better*. Students’ responses were coded as consistent with the experimental manipulation if the students ticked values above 5 when a downward comparison was induced, below 5 when an upward comparison was induced, or between 3 and 7 when a lateral comparison was induced. $N=411$

not substantially related to the frequency of responses consistent with the manipulation ($\Delta = 0.8\%$).

Although the students did not know their actual achievement at the time of the manipulation check (see Sect. 4.4), it is conceivable that they developed an impression of their actual achievement as they worked through the ability tests, which could have affected their achievement ratings in addition to the manipulated achievement feedback. Therefore, I conducted additional analyses investigating whether students’ ratings of their own achievement were more likely to fail the manipulation check, the more that students’ purported achievement differed from their actual achievement. However, there was little evidence, at best, for this assumption.

On the one hand, I found no significant relations between a “discrepancy variable”, which I calculated by summing the absolute values of the four differences between the corresponding purported and actual percentile ranks in the two ability tests at the two measurement points (the actual percentile ranks being calculated using the age-specific norms from the manual of the I-S-T 2000 R; Liepmann et al., 2007), and the two variables that indicated whether the students were included (coded as 1) or not included (coded as 0) in the analyses examining comparison effects on their FAT self-concept ($r = -.06$, $p = .20$) and VAT self-concept ($r = -.07$, $p = .17$). On the other hand, I mostly found no significant relations between the variables indicating whether students’ achievement ratings in social, dimensional, and temporal comparisons were consistent or inconsistent with the experimental manipulation and relevant achievement variables (see Table 3). Only students in the

Table 3 Correlations between successful manipulation and students' actual achievement differences

Triggered Comparison	FAT			VAT						
	<i>r</i>	<i>p</i>	95% CI	90% CI	<i>n</i>	<i>r</i>	<i>p</i>	95% CI	90% CI	<i>n</i>
<i>Social</i>										
Upward	.05	.68	[-.17, .25]	[-.13, .22]	87	.04	.70	[-.17, .24]	[-.13, .21]	92
Lateral	.01	.92	[-.12, .13]	[-.10, .11]	238	-.01	.92	[-.14, .12]	[-.12, .10]	232
Downward	.21	.05	[-.00, .41]	[.03, .38]	86	.25	.02	[.04, .44]	[.07, .41]	87
<i>Dimensional</i>										
Upward	-.04	.71	[-.25, .17]	[-.22, .14]	89	-.09	.41	[-.29, .12]	[-.26, .09]	93
Lateral	-.03	.66	[-.16, .10]	[-.14, .08]	229	-.06	.40	[-.18, .07]	[-.16, .05]	229
Downward	.06	.55	[-.14, .26]	[-.11, .23]	93	.02	.86	[-.19, .23]	[-.16, .19]	89
<i>Temporal</i>										
Upward	.01	.95	[-.27, .29]	[-.23, .24]	50	-.10	.48	[-.37, .18]	[-.33, .14]	50
Lateral	-.06	.26	[-.17, .05]	[-.16, .03]	315	.04	.46	[-.07, .15]	[-.05, .13]	315
Downward	.28	.06	[-.02, .53]	[.03, .49]	46	.31	.04	[.02, .55]	[.07, .51]	46

FAT = Figural Ability Test. VAT = Verbal Ability Test. The correlations refer to relations between different variables. In the case of a triggered social upward or social downward comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in social comparison was successful (coded as 1) or not (coded as 0) and students' mean achievement in the FAT/VAT (in percentile ranks) is shown. In the case of a triggered social lateral comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in social comparison was successful (coded as 1) or not (coded as 0) and the *absolute value* of the difference between students' mean achievement in the FAT/VAT (in percentile ranks) and the percentile rank of 50 is shown. In the case of a triggered dimensional upward or dimensional downward comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in dimensional comparison was successful (coded as 1) or not (coded as 0) and the difference between students' mean achievement in the FAT/VAT and students' mean achievement in the VAT/FAT (in percentile ranks) is shown. In the case of a triggered dimensional lateral comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in dimensional comparison was successful (coded as 1) or not (coded as 0) and the *absolute value* of the difference between students' mean achievement in the FAT and students' mean achievement in the VAT (in percentile ranks) is shown. In the case of a triggered temporal upward or temporal downward comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in temporal comparison was successful (coded as 1) or not (coded as 0) and the *absolute value* of the difference between students' mean achievement in the FAT/VAT and students' mean achievement in the VAT (in percentile ranks) is shown. In the case of a triggered temporal comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in temporal comparison was successful (coded as 1) or not (coded as 0) and the *absolute value* of the difference between students' mean achievement in the FAT/VAT and students' mean achievement in the VAT (in percentile ranks) is shown.

Table 3 (continued)

0) and the difference between students' achievement in the FAT/VAT at the second measurement point and students' achievement in the FAT/VAT at the first measurement point (in percentile ranks) is shown. In the case of a triggered temporal lateral comparison with respect to students' achievement in the FAT/VAT, the correlation between the variable indicating whether the manipulation of students' FAT/VAT achievement in temporal comparison was successful (coded as 1) or not (coded as 0) and the *absolute value* of the difference between students' achievement in the FAT/VAT at the second measurement point and students' achievement in the FAT/VAT at the first measurement point (in percentile ranks) is shown. The items used for the manipulation check asked the students to assess their own achievement in the FAT and VAT in comparison to the reference group ("Social"), to the other test ("Dimensional"), and at the second measurement point compared to the first measurement point ("Temporal"), on a Likert scale ranging from 1 = *much worse* to 9 = *much better*. Students' responses were coded as consistent with the experimental manipulation (i.e., the manipulation with regard to the specific variable was considered successful) if the students ticked values above 5 when a downward comparison was induced, below 5 when an upward comparison was induced, or between 3 and 7 when a lateral comparison was induced. Students' achievements in the FAT and VAT were calculated in three steps: First, the raw scores obtained by the students in the subtests of the FAT and VAT were converted into standard values according to the appropriate age norm for persons with a high school diploma, as stated in the manual of the I-S-T 2000 R (Liepmann et al., 2007). Second, the standard values obtained in the three subtests of the FAT and VAT, respectively, were averaged. Third, the averaged standard values were converted into percentile ranks. The mean achievement in the FAT and VAT corresponds to the average achievement in the respective test from the two measurement points. $N=411$

experimental conditions that triggered social downward comparisons (i.e., Conditions 2 and 5 or 7) were more likely to rate their achievement in one test in social comparison consistently with the manipulation (i.e., to tick values above 5) the better they actually performed in the respective test, and students in the experimental condition that triggered temporal downward comparisons (i.e., Condition 8) were more likely to rate their achievement in one test in temporal comparison consistently with the manipulation (i.e., to tick values above 5) the more they actually improved (or the less they actually worsened) in the respective test (however, for the FAT, these relations were only statistically significant at the 10% significance level).

4.4 Procedure

The study was conducted online via web conferencing in BigBlueButton between November 2020 and July 2022. Students participated in small groups. Each session lasted almost two hours and was moderated by an experimenter who was available to answer students' questions at any time. When registering for the study, students selected two participation dates about two weeks apart. They were then sent a link to the BigBlueButton room. Data were collected via Limesurvey questionnaires. The links to these questionnaires were sent within the BigBlueButton sessions. Individual identification numbers were used to merge the data from both measurement points.

At the beginning of the first session, students were told as a cover story that the goal of the study was to test digital versions of figural and verbal ability tests that had been used for several years in paper-and-pencil versions. They were also informed that the figural and verbal tests existed in two parallel forms that were identical in terms of task type and difficulty. In each session, they would complete one parallel form of each test. At the end of the second session, they would receive feedback on their achievement in both tests from both measurement points. This feedback would be generated automatically so that neither the examiner nor the other participants would know the individual results. Moreover, the students would receive feedback about the achievement of another (anonymous) student of the group.

Following this introduction, the students were asked to answer some demographic questions. Subsequently, the experimenter presented them with sample items from the FAT and VAT subtests that were solved as a group practice exercise. After all questions about the subtests were clarified, the students were asked to rate their perceived ability to solve FAT and VAT tasks, based on the sample items presented. Then, the students worked on the six subtests of the first parallel version. In line with the test manual, they had between 6 and 10 min to complete each subtest. A countdown onscreen showed the time remaining. Before the start of each subtest, the respective instructions were repeated within a 30-s break. After having completed all subtests, the students worked on some items that were not relevant for the present study.

At the beginning of the second session, the different task types of the FAT and VAT were briefly revised. Subsequently, the students worked on the six subtests of the second parallel version (similar to T1). After that, the experimenter prepared the

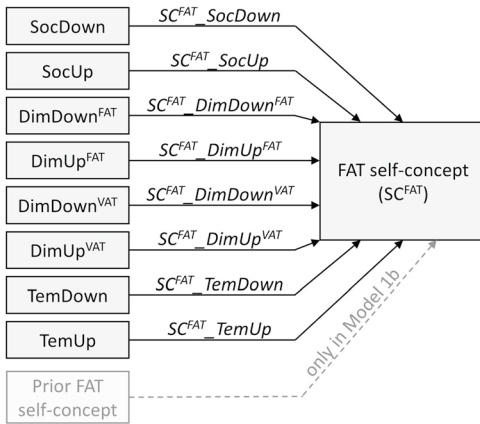
students for the achievement feedback by explaining how to interpret a percentile rank and interpreting some arbitrary percentile ranks together with the students as a group. As soon as all students had understood how to interpret a percentile rank, they successively received the manipulated feedback on their own achievement and that of the peer. Following each feedback occasion, they were asked to assess their own or the peer's ability to solve FAT and VAT tasks (the peer's ability from their own and the peer's perspective) and to rate their own or the peer's achievement on the two ability tests, respectively. During these rating processes the relevant percentile ranks were shown. After all students had assessed both their own and the peer's abilities and achievements, they were informed about the actual purpose of the study. They were also told that the received achievement feedback had been manipulated, and finally received their true results in the ability tests.

The order in which students worked on the FAT and VAT and the test-specific items, as well as the order in which the students received feedback about their own and the peer's achievement, varied randomly between the groups. Furthermore, it was randomized whether the students should assess the peer's abilities first from their own or from the peer's perspective. All items of one scale were also presented in randomized order.

4.5 Statistical analyses

I tested my hypotheses using regression analyses in Mplus 7.4 (Muthén & Muthén, 2015). For model estimation, I used the robust maximum likelihood estimator (MLR). Overall, I specified six models, in which students' ability assessments (at T2) were regressed on eight dummy variables, indicating the experimental Conditions 2–9 (i.e., Condition 1 served as the reference condition). Figures 2, 3, 4 and 5 illustrate these different models. As shown, Model 1a and Model 1b included students' own FAT self-concept as criterion. These models only differed from each other in that in Model 1b I also controlled for students' FAT self-concept at T1. Similarly, Model 2a and Model 2b included students' own VAT self-concept as criterion and they differed from each other only in that in Model 2b students' VAT self-concept at T1 was also controlled for. Model 3 included two criteria: students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks. Thus, this model allowed me to test for significant differences in the strength of the same comparison effects on the two different kinds of ability ratings, made by the same participants based on the same information about the peer's achievement. Similarly, Model 4 included both students' inferred VAT self-concept of a peer and students' own assessment of a peer's ability to solve VAT tasks as criteria.

To calculate the comparison effects as defined in Sect. 4.1, I used the Model Constraint option implemented in Mplus to subtract the effects of the conditions that triggered downward and upward comparisons of the same comparison type (see Figs. 2, 3, 4 and 5). Thus, positive differences of the two effects subtracted from



Calculation of the comparison effects:

Social comparison effect:

$$SC^{FAT}_{SocDown} - SC^{FAT}_{SocUp}$$

→ Hypothesis 1.1 (+)

Dimensional comparison effect:

$$SC^{FAT}_{DimDown^{FAT}} - SC^{FAT}_{DimUp^{FAT}}$$

→ Hypothesis 2.1 (+)

Temporal comparison effect:

$$SC^{FAT}_{TemDown} - SC^{FAT}_{TemUp}$$

→ Hypothesis 3.1 (+)

Comparison of the comparison effects:

Social – dimensional comparison effect:

$$|SC^{FAT}_{SocDown} - SC^{FAT}_{SocUp}| -$$

$$|SC^{FAT}_{DimDown^{FAT}} - SC^{FAT}_{DimUp^{FAT}}|$$

→ Hypothesis 4.1 (+)

Social – temporal comparison effect:

$$|SC^{FAT}_{SocDown} - SC^{FAT}_{SocUp}| -$$

$$|SC^{FAT}_{TemDown} - SC^{FAT}_{TemUp}|$$

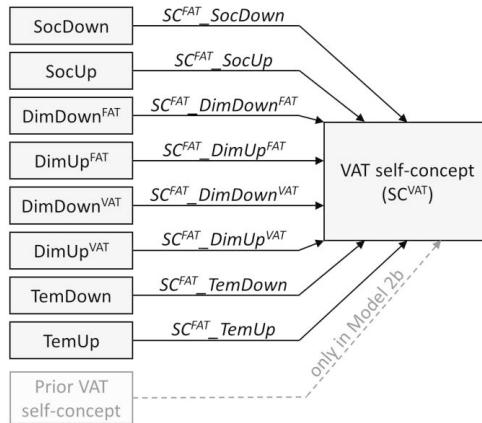
→ Hypothesis 5.1 (+)

Dimensional – temporal comparison effect:

$$|SC^{FAT}_{DimDown^{FAT}} - SC^{FAT}_{DimUp^{FAT}}| -$$

$$|SC^{FAT}_{TemDown} - SC^{FAT}_{TemUp}|$$

Fig. 2 Model 1a und Model 1b: prediction of students’ own FAT self-concept before (Model 1a) and after (Model 1b) controlling for prior FAT self-concept. FAT=Figural Ability Test. VAT=Verbal Ability Test. Students’ FAT self-concept used as criterion was assessed at the second measurement point (T2). Students’ FAT self-concept used as predictor in Model 1b was assessed at the first measurement point (T1). The “+” after the hypotheses indicates that the respective effects are assumed to be positive



Calculation of the comparison effects:

Social comparison effect:

$$SC^{VAT}_{SocDown} - SC^{VAT}_{SocUp}$$

→ Hypothesis 1.1 (+)

Dimensional comparison effect:

$$SC^{VAT}_{DimDown^{VAT}} - SC^{VAT}_{DimUp^{VAT}}$$

→ Hypothesis 2.1 (+)

Temporal comparison effect:

$$SC^{VAT}_{TemDown} - SC^{VAT}_{TemUp}$$

→ Hypothesis 3.1 (+)

Comparison of the comparison effects:

Social – dimensional comparison effect:

$$|SC^{VAT}_{SocDown} - SC^{VAT}_{SocUp}| - |SC^{VAT}_{DimDown^{VAT}} - SC^{VAT}_{DimUp^{VAT}}|$$

→ Hypothesis 4.1 (+)

Social – temporal comparison effect:

$$|SC^{VAT}_{SocDown} - SC^{VAT}_{SocUp}| - |SC^{VAT}_{TemDown} - SC^{VAT}_{TemUp}|$$

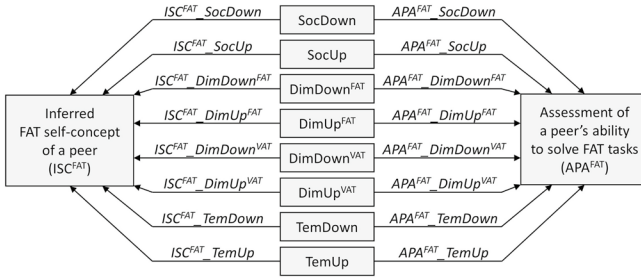
→ Hypothesis 5.1 (+)

Dimensional – temporal comparison effect:

$$|SC^{VAT}_{DimDown^{VAT}} - SC^{VAT}_{DimUp^{VAT}}| - |SC^{VAT}_{TemDown} - SC^{VAT}_{TemUp}|$$

Fig. 3 Model 2a und Model 2b: prediction of students’ own VAT self-concept before (Model 2a) and after (Model 2b) controlling for prior VAT self-concept. FAT=Figural Ability Test. VAT=Verbal Ability Test. Students’ VAT self-concept used as criterion was assessed at the second measurement point (T2). Students’ VAT self-concept used as predictor in Model 1b was assessed at the first measurement point (T1). The “+” after the hypotheses indicates that the respective effects are assumed to be positive

each other would indicate significantly higher ability ratings after downward comparisons as opposed to upward comparisons, and thus significant comparison effects, as predicted in Hypotheses 1–3. Nonetheless, to further understand the emergence of potential comparison effects, I also looked at the effects of the dummy variables,



Calculation of the comparison effects on students' inferred self-concept of a peer:

Social comparison effect:

$ISC^{FAT}_{SocDown} - ISC^{FAT}_{SocUp}$
 → Hypothesis 1.2 (+)

Dimensional comparison effect:

$ISC^{FAT}_{DimDown^{FAT}} - ISC^{FAT}_{DimUp^{FAT}}$
 → Hypothesis 2.2 (+)

Temporal comparison effect:

$ISC^{FAT}_{TemDown} - ISC^{FAT}_{TemUp}$
 → Hypothesis 3.2 (+)

Comparison of the comparison effects on students' inferred self-concept of a peer:

Social – dimensional comparison effect:

$|ISC^{FAT}_{SocDown} - ISC^{FAT}_{SocUp}| - |ISC^{FAT}_{DimDown^{FAT}} - ISC^{FAT}_{DimUp^{FAT}}|$
 → Hypothesis 4.2 (+)

Social – temporal comparison effect:

$|ISC^{FAT}_{SocDown} - ISC^{FAT}_{SocUp}| - |ISC^{FAT}_{TemDown} - ISC^{FAT}_{TemUp}|$
 → Hypothesis 5.2 (+)

Dimensional – temporal comparison effect:

$|ISC^{FAT}_{DimDown^{FAT}} - ISC^{FAT}_{DimUp^{FAT}}| - |ISC^{FAT}_{TemDown} - ISC^{FAT}_{TemUp}|$

Calculation of the comparison effects on students' own assessment of a peer's ability:

Social comparison effect:

$APA^{FAT}_{SocDown} - APA^{FAT}_{SocUp}$
 → Hypothesis 1.3 (+)

Dimensional comparison effect:

$APA^{FAT}_{DimDown^{FAT}} - APA^{FAT}_{DimUp^{FAT}}$

Temporal comparison effect:

$APA^{FAT}_{TemDown} - APA^{FAT}_{TemUp}$

Comparison of the comparison effects on students' own assessment of a peer's ability:

Social – dimensional comparison effect:

$|APA^{FAT}_{SocDown} - APA^{FAT}_{SocUp}| - |APA^{FAT}_{DimDown^{FAT}} - APA^{FAT}_{DimUp^{FAT}}|$
 → Hypothesis 4.3 (+)

Social – temporal comparison effect:

$|APA^{FAT}_{SocDown} - APA^{FAT}_{SocUp}| - |APA^{FAT}_{TemDown} - APA^{FAT}_{TemUp}|$
 → Hypothesis 5.3 (+)

Dimensional – temporal comparison effect:

$|APA^{FAT}_{DimDown^{FAT}} - APA^{FAT}_{DimUp^{FAT}}| - |APA^{FAT}_{TemDown} - APA^{FAT}_{TemUp}|$

Comparison of the comparison effects on students' inferred self-concept of a peer and students' own assessment of a peer's ability :

Social – social comparison effect:

$|ISC^{FAT}_{SocDown} - ISC^{FAT}_{SocUp}| - |APA^{FAT}_{SocDown} - APA^{FAT}_{SocUp}|$

Dimensional – dimensional comparison effect:

$|ISC^{FAT}_{DimDown^{FAT}} - ISC^{FAT}_{DimUp^{FAT}}| - |APA^{FAT}_{DimDown^{FAT}} - APA^{FAT}_{DimUp^{FAT}}|$

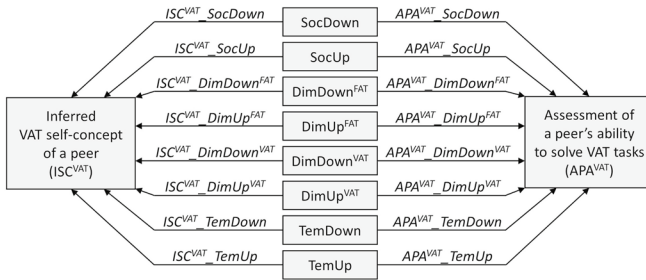
Temporal – temporal comparison effect:

$|ISC^{FAT}_{TemDown} - ISC^{FAT}_{TemUp}| - |APA^{FAT}_{TemDown} - APA^{FAT}_{TemUp}|$

Fig. 4 Model 3: prediction of students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks. FAT=Figural Ability Test. VAT=Verbal Ability Test. The criteria (i.e., students' inferred FAT self-concept of a peer and students' assessment of a peer's ability to solve FAT tasks) were assessed at the second measurement point (T2). The "+" after the hypotheses indicates that the respective effects are assumed to be positive

which indicated differences in the ability ratings after downward or upward comparisons in relation to lateral comparisons.

I also used the Model Constraint option to compare the strength of the social, dimensional, and temporal comparison effects on the same ability ratings. This



Calculation of the comparison effects on students' inferred self-concept of a peer:

Social comparison effect:

$$ISC^{VAT}_{SocDown} - ISC^{VAT}_{SocUp}$$

→ Hypothesis 1.2 (+)

Dimensional comparison effect:

$$ISC^{VAT}_{DimDown^{VAT}} - ISC^{VAT}_{DimUp^{VAT}}$$

→ Hypothesis 2.2 (+)

Temporal comparison effect:

$$ISC^{VAT}_{TemDown} - ISC^{VAT}_{TemUp}$$

→ Hypothesis 3.2 (+)

Comparison of the comparison effects on students' inferred self-concept of a peer:

Social – dimensional comparison effect:

$$|ISC^{VAT}_{SocDown} - ISC^{VAT}_{SocUp}| - |ISC^{VAT}_{DimDown^{VAT}} - ISC^{VAT}_{DimUp^{VAT}}|$$

→ Hypothesis 4.2 (+)

Social – temporal comparison effect:

$$|ISC^{VAT}_{SocDown} - ISC^{VAT}_{SocUp}| - |ISC^{VAT}_{TemDown} - ISC^{VAT}_{TemUp}|$$

→ Hypothesis 5.2 (+)

Dimensional – temporal comparison effect:

$$|ISC^{VAT}_{DimDown^{VAT}} - ISC^{VAT}_{DimUp^{VAT}}| - |ISC^{VAT}_{TemDown} - ISC^{VAT}_{TemUp}|$$

Calculation of the comparison effects on students' own assessment of a peer's ability:

Social comparison effect:

$$APA^{VAT}_{SocDown} - APA^{VAT}_{SocUp}$$

→ Hypothesis 1.3 (+)

Dimensional comparison effect:

$$APA^{VAT}_{DimDown^{VAT}} - APA^{VAT}_{DimUp^{VAT}}$$

Temporal comparison effect:

$$APA^{VAT}_{TemDown} - APA^{VAT}_{TemUp}$$

Comparison of the comparison effects on students' own assessment of a peer's ability:

Social – dimensional comparison effect:

$$|APA^{VAT}_{SocDown} - APA^{VAT}_{SocUp}| - |APA^{VAT}_{DimDown^{VAT}} - APA^{VAT}_{DimUp^{VAT}}|$$

→ Hypothesis 4.3 (+)

Social – temporal comparison effect:

$$|APA^{VAT}_{SocDown} - APA^{VAT}_{SocUp}| - |APA^{VAT}_{TemDown} - APA^{VAT}_{TemUp}|$$

→ Hypothesis 5.3 (+)

Dimensional – temporal comparison effect:

$$|APA^{VAT}_{DimDown^{VAT}} - APA^{VAT}_{DimUp^{VAT}}| - |APA^{VAT}_{TemDown} - APA^{VAT}_{TemUp}|$$

Comparison of the comparison effects on students' inferred self-concept of a peer and students' own assessment of a peer's ability :

Social – social comparison effect:

$$|ISC^{VAT}_{SocDown} - ISC^{VAT}_{SocUp}| - |APA^{VAT}_{SocDown} - APA^{VAT}_{SocUp}|$$

Dimensional – dimensional comparison effect:

$$|ISC^{VAT}_{DimDown^{VAT}} - ISC^{VAT}_{DimUp^{VAT}}| - |APA^{VAT}_{DimDown^{VAT}} - APA^{VAT}_{DimUp^{VAT}}|$$

Temporal – temporal comparison effect:

$$|ISC^{VAT}_{TemDown} - ISC^{VAT}_{TemUp}| - |APA^{VAT}_{TemDown} - APA^{VAT}_{TemUp}|$$

Fig. 5 Model 4: prediction of students' inferred VAT self-concept of a peer and students' own assessment of a peer's ability to solve VAT tasks. FAT=Figural Ability Test. VAT=Verbal Ability Test. The criteria (i.e., students' inferred VAT self-concept of a peer and students' assessment of a peer's ability to solve VAT tasks) were assessed at the second measurement point (T2). The “+” after the hypotheses indicates that the respective effects are assumed to be positive

allowed me to test Hypotheses 4 and 5. Moreover, I used the Model Constraint option in Model 3 and Model 4 to compare the strength of the effects of the same comparison types on students' inferred self-concept of a peer and their own

assessment of a peer's ability. When comparing the strength of different comparison effects, I referred to the absolute values of the comparison effects. I tested all hypothesized effects one-sided (including effects of the dummy variables that contributed to the comparison effects in the expected direction). All other effects were tested two-sided.

I specified the ability ratings as manifest variables to enhance the comparability of the findings across the different models and with the descriptive statistics. In line with that, the three prior experimental studies examining the joint effects of social, dimensional, and temporal comparisons on students' self-concepts (Wolff et al., 2018b; Zell & Strickhouser, 2020) had also specified students' self-concepts as manifest variables. Nevertheless, I note that all my main findings (i.e., all findings concerning the significance of the comparison effects and the comparison of the comparison effects) were also replicated in additional analyses, in which the ability ratings were specified as latent variables. The results of these analyses can be found in Tables S4–S7 in the Supplemental Material. Moreover, the Supplemental Material sets out the syntaxes used to calculate the models in Mplus.

5 Results

5.1 Preliminary analyses

Table 4 shows the means and standard deviations of the ability assessments in the different experimental conditions. Since students were randomly assigned to the experimental conditions, the mean self-concepts at T1 should not differ between the groups. However, although one-way ANOVAs revealed no significant differences between the groups regarding students' initial FAT self-concept, $F(8, 250) = 1.96$, $p \geq .05$, and initial VAT self-concept, $F(8, 266) = 0.50$, $p = .86$, post-hoc tests indicated that students in the DimDown^{FAT} condition had a significantly higher initial FAT self-concept than students in the Reference ($\Delta M = 0.57$, $p = .04$), DimUp^{FAT} ($\Delta M = 0.85$, $p < .01$), DimDown^{VAT} ($\Delta M = 0.74$, $p = .01$), DimUp^{VAT} ($\Delta M = 0.90$, $p < .01$), and TemDown ($\Delta M = 0.70$, $p = .01$) conditions. Moreover, students in the SocDown condition had a significantly higher initial FAT self-concept than students in the DimUp^{FAT} ($\Delta M = 0.55$, $p = .04$) and DimUp^{VAT} ($\Delta M = 0.60$, $p = .04$) conditions (all other $\Delta |M| \leq 0.47$, all $p \geq .10$). For the investigation of comparison effects, the difference between the DimDown^{FAT} and DimUp^{FAT} conditions was somewhat problematic, as these two conditions were compared with each other to examine the dimensional comparison effect on students' FAT self-concept. Accordingly, it was reasonable to conduct an additional analysis in which students' initial FAT self-concept was controlled for (i.e., Model 1b) to test whether a potential dimensional comparison effect on students' FAT self-concept at T2 resulted from differences in their FAT self-concept at T1 rather than from the experimental manipulation.

Table 4 Means and standard deviations of ability assessments and sample sizes in the different experimental conditions

Condition	Students' Own Self-Concept (T1)						Students' Own Self-Concept (T2)						Student's Inferred Self-Concept of a Peer (T2)						Students' Own Assessment of a Peer's Ability (T2)					
	FAT		VAT		FAT		VAT		FAT		VAT		FAT		VAT		FAT		VAT		FAT		VAT	
	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N
1 ("Reference")	4.20	1.02	40	4.88	0.76	42	3.70	0.76	40	4.19	0.63	42	4.08	0.74	40	4.27	0.67	37	4.33	0.70	40	4.35	0.70	37
2 ("SocDown")	4.47	1.17	25	4.90	0.63	29	4.39	0.98	25	4.72	0.67	29	4.60	0.77	45	4.81	0.64	45	4.95	0.72	45	5.12	0.51	45
3 ("SocUp")	4.33	0.79	26	4.62	0.78	29	3.11	0.83	26	3.47	0.78	29	2.78	0.55	32	3.12	0.74	31	2.86	0.72	32	2.96	0.75	31
4 ("DimDown" ^{FAT})	4.77	0.98	21	4.73	0.70	23	3.97	0.72	21	3.59	0.64	23	4.33	0.50	21	3.15	0.72	19	4.22	0.69	21	3.04	0.56	19
5 ("DimUp" ^{FAT})	3.92	0.90	32	4.86	0.73	36	3.55	1.04	32	4.81	0.61	36	3.41	0.90	33	4.95	0.55	40	3.71	0.82	33	5.21	0.39	40
6 ("DimDown" ^{VAT})	4.03	0.99	25	4.76	0.82	31	2.83	0.93	25	4.22	0.67	31	3.01	0.84	34	4.28	0.60	34	3.13	0.74	34	4.40	0.56	34
7 ("DimUp" ^{VAT})	3.87	1.20	24	4.74	0.79	23	3.95	1.03	24	3.89	0.57	23	4.72	0.69	30	3.88	0.72	23	4.95	0.72	30	4.06	0.64	23
8 ("TemDown")	4.07	1.07	34	4.69	0.80	33	3.51	0.91	34	4.06	0.73	33	3.68	0.63	33	4.00	0.61	35	4.00	0.66	33	4.15	0.59	35
9 ("TemUp")	4.32	0.93	32	4.80	0.65	29	3.54	0.72	32	3.99	0.61	29	3.99	0.59	29	4.04	0.61	29	4.19	0.51	29	4.10	0.57	29

FAT = Figural Ability Test. VAT = Verbal Ability Test. T1 = First measurement point. T2 = Second measurement point

5.2 Comparison effects on students' own self-concepts

Table 5 presents the results of the regression analyses predicting students' FAT self-concept (Model 1a and Model 1b). Table 6 shows the results of the regression analyses predicting students' VAT self-concept (Model 2a and Model 2b).

In line with Hypothesis 1.1, there were significant social comparison effects both before and after controlling for prior self-concepts, indicating higher self-concepts following social downward comparisons and lower self-concepts following social upward comparisons (all $1.11 \leq B \leq 1.28$, all $p < .001$, all $0.38 \leq \beta \leq 0.49$). As predicted in Hypothesis 2.1, the dimensional comparison effects were significant in the analyses that did not control for prior self-concepts, indicating higher self-concepts following dimensional downward comparisons and lower self-concepts following dimensional upward comparisons (all $0.33 \leq B \leq 0.42$, all $p \leq .08$, all $0.12 \leq \beta \leq 0.13$). However, after controlling for prior self-concepts, only the dimensional comparison effect on students' VAT self-concept remained significant ($B = 0.32$, $p = .03$, $\beta = 0.12$), whereas the dimensional comparison effect on students' FAT self-concept lost significance ($B = -0.06$, $p = .78$). In contrast to Hypothesis 3.1, none of the temporal comparison effects were significant, either before or after controlling for prior self-concepts (all $0.12 \leq |B|$, all $p \geq .35$).

The comparisons of the different kinds of comparison effects revealed that the social comparison effects were significantly stronger than the dimensional comparison effects in all four analyses (all $0.79 \leq B \leq 1.15$, all $p \leq .01$, all $0.27 \leq \beta \leq 0.38$), which supported Hypothesis 4.1. Similarly, and in accord with Hypothesis 5.1, in all four analyses the social comparison effects were significantly stronger than the temporal comparison effects (all $0.99 \leq B \leq 1.26$, all $p < .001$, all $0.34 \leq \beta \leq 0.47$). The dimensional and temporal comparison effects did not differ from each other in respect to strength in any of the analyses (all $0.39 \leq |B|$, all $p \geq .20$).

An exploratory inspection of the effects of the dummy variables showed that in all four analyses students' self-concepts were significantly higher in the SocDown condition than in the Reference condition (all $0.52 \leq B \leq 0.70$, all $p < .01$, all $0.17 \leq \beta \leq 0.21$), whereas they were significantly lower in the SocUp condition than in the Reference condition (all $-0.72 \leq B \leq -0.59$, all $p < .01$, all $-0.29 \leq \beta \leq -0.18$). Students' FAT self-concept did not differ significantly between the Reference condition and the DimDown^{FAT} condition or between the Reference condition and the DimUp^{FAT} condition (all $0.27 \leq |B|$, all $p \geq .16$). Their VAT self-concept did not differ significantly between the Reference condition and the DimDown^{VAT} condition (all $0.09 \leq |B|$, all $p \geq .50$), but was significantly lower in the DimUp^{VAT} condition compared to the Reference condition (all $-0.30 \leq B \leq -0.23$, all $p \leq .05$, all $-0.11 \leq \beta \leq -0.08$). The differences between students' self-concepts in the Reference condition and the TemDown and TemUp conditions were nonsignificant (all $0.20 \leq |B|$, all $p \geq .18$), except for a significantly lower FAT self-concept in the TemUp condition after controlling for prior FAT self-concept ($B = -0.23$, $p = .08$, $\beta = -0.08$).

Table 5 Results of the regression analyses predicting students' own FAT self-concept (Model 1a and Model 1b)

Predictor	FAT Self-Concept (T1)				FAT Self-Concept (T2)							
	Estimate	S.E	p	95% CI	90% CI	β	Estimate	S.E	p	95% CI	90% CI	β
Intercept	3.70	0.12	.00	[3.46, 3.93]	[3.50, 3.89]	3.85	1.37	0.21	.00	[0.96, 1.78]	[1.03, 1.71]	1.43
SocDown	0.70	0.23	.00	[0.25, 1.14]	[0.33, 1.07]	0.21	0.55	0.14	.00	[0.27, 0.83]	[0.31, 0.79]	0.17
SocUp	-0.59	0.20	.00	[-0.98, -0.20]	[-0.92, -0.26]	-0.18	-0.66	0.15	.00	[-0.96, -0.37]	[-0.91, -0.41]	-0.21
DimDown ^{FAT}	0.27	0.19	.16	[-0.11, 0.65]	[-0.05, 0.59]	0.08	-0.04	0.17	.80	[-0.38, 0.29]	[-0.33, 0.24]	-0.01
DimUp ^{FAT}	-0.14	0.22	.51	[-0.57, 0.28]	[-0.50, 0.21]	-0.05	0.01	0.16	.94	[-0.31, 0.33]	[-0.26, 0.28]	0.01
DimDown ^{VAT}	-0.87	0.22	.00	[-1.30, -0.44]	[-1.23, -0.51]	-0.27	-0.78	0.20	.00	[-1.17, -0.39]	[-1.11, -0.45]	-0.24
DimUp ^{VAT}	0.26	0.24	.28	[-0.21, 0.72]	[-0.14, 0.65]	0.08	0.44	0.17	.01	[0.11, 0.77]	[0.16, 0.72]	0.13
TemDown	-0.18	0.19	.35	[-0.56, 0.20]	[-0.50, 0.14]	-0.06	-0.11	0.15	.45	[-0.40, 0.18]	[-0.36, 0.13]	-0.04
TemUp	-0.16	0.17	.36	[-0.50, 0.18]	[-0.44, 0.13]	-0.06	-0.23	0.13	.08	[-0.48, 0.03]	[-0.44, -0.01]	-0.08
FAT Self-Concept (T1)	—	—	—	—	—	—	0.55	0.05	.00	[0.47, 0.64]	[0.48, 0.63]	0.59
<i>Model Constraint</i>												
Social Comparison Effect	1.28	0.25	.00	[0.79, 1.78]	[0.87, 1.70]	0.40	1.21	0.17	.00	[0.88, 1.54]	[0.94, 1.49]	0.38
Dimensional Comparison Effect	0.42	0.24	.08	[-0.05, 0.88]	[0.03, 0.81]	0.13	-0.06	0.20	.78	[-0.45, 0.34]	[-0.39, 0.28]	-0.02
Temporal Comparison Effect	-0.02	0.20	.91	[-0.41, 0.37]	[-0.35, 0.30]	-0.01	0.12	0.15	.45	[-0.19, 0.42]	[-0.14, 0.37]	0.04
Social – Dimensional Comparison Effect	0.87	0.35	.01	[0.19, 1.54]	[0.30, 1.44]	0.27	1.15	0.26	.00	[0.64, 1.67]	[0.72, 1.59]	0.36
Social – Temporal Comparison Effect	1.26	0.32	.00	[0.64, 1.89]	[0.74, 1.79]	0.39	1.09	0.23	.00	[0.65, 1.54]	[0.72, 1.47]	0.34
Dimensional – Temporal Comparison Effect	0.39	0.31	.20	[-0.21, 1.00]	[-0.11, 0.90]	0.12	-0.06	0.26	.81	[-0.56, 0.44]	[-0.48, 0.36]	-0.02

FAT = Figural Ability Test. VAT = Verbal Ability Test. T1 = First measurement point. T2 = Second measurement point. The specification of the effects under “Model Constraint” can be found in Fig. 2. N=259

Table 6 Results of the regression analyses predicting students' own VAT self-concept (Model 2a and Model 2b)

Predictor	VAT Self-Concept (T2)				VAT Self-Concept (T2)							
	Estimate	S.E	p	95% CI	90% CI	β	Estimate	S.E	p	95% CI	90% CI	β
Intercept	4.19	0.10	.00	[4.00, 4.38]	[4.03, 4.35]	5.41	1.73	0.24	.00	[1.26, 2.20]	[1.34, 2.13]	2.24
SocDown	0.53	0.16	.00	[0.22, 0.83]	[0.27, 0.78]	0.21	0.52	0.12	.00	[0.29, 0.75]	[0.33, 0.72]	0.21
SocUp	-0.72	0.17	.00	[-1.06, -0.38]	[-1.00, -0.44]	-0.29	-0.59	0.15	.00	[-0.87, -0.30]	[-0.83, -0.35]	-0.23
DimDown ^{FAT}	-0.60	0.16	.00	[-0.92, -0.29]	[-0.87, -0.34]	-0.22	-0.53	0.15	.00	[-0.82, -0.23]	[-0.78, -0.28]	-0.19
DimUp ^{FAT}	0.62	0.14	.00	[0.34, 0.89]	[0.39, 0.84]	0.27	0.63	0.11	.00	[0.41, 0.84]	[0.45, 0.81]	0.27
DimDown ^{VAT}	0.03	0.15	.85	[-0.27, 0.33]	[-0.22, 0.28]	0.01	0.09	0.14	.50	[-0.18, 0.37]	[-0.14, 0.32]	0.04
DimUp ^{VAT}	-0.30	0.15	.05	[-0.59, -0.00]	[-0.55, -0.05]	-0.11	-0.23	0.12	.05	[-0.45, 0.00]	[-0.42, -0.03]	-0.08
TemDown	-0.13	0.16	.41	[-0.44, 0.18]	[-0.39, 0.13]	-0.06	-0.03	0.12	.79	[-0.27, 0.21]	[-0.24, 0.17]	-0.01
TemUp	-0.20	0.15	.18	[-0.49, 0.09]	[-0.44, 0.05]	-0.08	-0.15	0.12	.19	[-0.38, 0.07]	[-0.34, 0.04]	-0.06
VAT Self-Concept (T1)	—	—	—	—	—	—	0.50	0.05	.00	[0.41, 0.60]	[0.43, 0.58]	0.48
<i>Model Constraint</i>												
Social Comparison Effect	1.25	0.19	.00	[0.88, 1.62]	[0.94, 1.56]	0.49	1.11	0.15	.00	[0.81, 1.41]	[0.86, 1.36]	0.44
Dimensional Comparison Effect	0.33	0.17	.05	[0.00, 0.66]	[0.06, 0.60]	0.12	0.32	0.15	.03	[0.03, 0.61]	[0.08, 0.56]	0.12
Temporal Comparison Effect	0.07	0.17	.69	[-0.26, 0.40]	[-0.21, 0.34]	0.02	0.12	0.13	.35	[-0.13, 0.37]	[-0.09, 0.33]	0.05
Social – Dimensional Comparison Effect	0.92	0.25	.00	[0.43, 1.41]	[0.51, 1.33]	0.38	0.79	0.21	.00	[0.38, 1.20]	[0.44, 1.13]	0.32
Social – Temporal Comparison Effect	1.18	0.25	.00	[0.69, 1.68]	[0.77, 1.60]	0.47	0.99	0.20	.00	[0.60, 1.38]	[0.66, 1.32]	0.39
Dimensional – Temporal Comparison Effect	0.26	0.24	.27	[-0.20, 0.73]	[-0.13, 0.65]	0.10	0.20	0.20	.31	[-0.18, 0.58]	[-0.12, 0.52]	0.07

FAT = Figural Ability Test. VAT = Verbal Ability Test. T1 = First measurement point. T2 = Second measurement point. The specification of the effects under "Model Constraint" can be found in Fig. 3. N = 275

5.3 Comparison effects on students' inferred self-concepts of a peer and students' own assessments of a peer's ability

Table 7 shows the results of the regression analyses predicting students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks (Model 3). Table 8 presents the results of the regression analyses predicting students' inferred VAT self-concept of a peer and students' own assessment of a peer's ability to solve VAT tasks (Model 4).

In line with Hypothesis 1.2, there were significant social comparison effects on students' inferred self-concepts of a peer, indicating higher inferred self-concepts following social downward comparisons and lower inferred self-concepts following social upward comparisons (all $1.69 \leq B \leq 1.82$, all $p < .001$, all $0.62 \leq \beta \leq 0.63$). Moreover, and as predicted in Hypothesis 1.3, there were significant social comparison effects on students' own assessments of a peer's ability, indicating higher ability assessments following social downward comparisons and lower ability assessments following social upward comparisons (all $2.09 \leq B \leq 2.16$, all $p < .001$, all $0.69 \leq \beta \leq 0.76$). Hypothesis 2.2 also found support: There were significant dimensional comparison effects on students' inferred self-concepts of a peer, indicating higher inferred self-concepts following dimensional downward comparisons and lower inferred self-concepts following dimensional upward comparisons (all $0.40 \leq B \leq 0.92$, all $p \leq .03$, all $0.13 \leq \beta \leq 0.29$). Similarly, there were significant dimensional comparison effects on students' own assessments of a peer's ability, indicating higher ability assessments following dimensional downward comparisons and lower ability assessments following dimensional upward comparisons (all $0.34 \leq B \leq 0.51$, all $p \leq .03$, all $0.10 \leq \beta \leq 0.17$). However, as opposed to Hypothesis 3.2, the temporal comparison effect on students' inferred FAT self-concept was significant, but indicated a higher inferred FAT self-concept following a temporal upward comparison and a lower inferred FAT self-concept following a temporal downward comparison ($B = -0.32$, $p = .04$, $\beta = -0.11$), while the temporal comparison effect on students' inferred VAT self-concept was nonsignificant ($B = -0.05$, $p = .77$). The temporal comparison effects on students' own assessments of a peer's ability to solve FAT or VAT tasks were also nonsignificant (all $0.19 \leq |B|$, all $p \geq .20$).

The comparisons of the different kinds of comparison effects on the same ability rating revealed that the social comparison effects were significantly stronger than the dimensional comparison effects, both on students' inferred self-concepts of a peer (all $0.90 \leq B \leq 1.29$, all $p \leq .001$, all $0.33 \leq \beta \leq 0.51$), which supported Hypothesis 4.2, and on their own assessments of a peer's ability (all $1.58 \leq B \leq 1.82$, all $p \leq .001$, all $0.52 \leq \beta \leq 0.66$), which supported Hypothesis 4.3. Furthermore, the social comparison effects were significantly stronger than the temporal comparison effects, both on students' inferred self-concepts of a peer (all $1.50 \leq B \leq 1.64$, all $p \leq .001$, all $0.51 \leq \beta \leq 0.61$), which supported Hypothesis 5.2, and on their own assessments of a peer's ability (all $1.90 \leq B \leq 2.11$, all $p \leq .001$, all $0.63 \leq \beta \leq 0.75$), which supported Hypothesis 5.3. The dimensional and temporal comparison effects mostly did not differ from each other in terms of strength (all $0.36 \leq |B|$, all $p \geq .13$). Only the dimensional comparison effect on students' inferred FAT self-concept of a peer was

Table 7 Results of the regression analyses predicting students' inferred FAT self-concept of a peer and student's own assessment of a peer's ability to solve FAT tasks (Model 3)

Predictor	Inferred FAT Self-Concept of a Peer				Assessment of a Peer's Ability to Solve FAT Tasks							
	Estimate	S.E	p	95% CI	90% CI	β	Estimate	S.E	p	95% CI	90% CI	β
Intercept	4.08	0.12	.00	[3.85, 4.31]	[3.89, 4.27]	4.28	4.33	0.11	.00	[4.11, 4.54]	[4.14, 4.51]	4.40
SocDown	0.52	0.16	.00	[0.20, 0.83]	[0.25, 0.78]	0.19	0.63	0.15	.00	[0.33, 0.93]	[0.38, 0.88]	0.23
SocUp	-1.30	0.15	.00	[-1.60, -1.01]	[-1.55, -1.06]	-0.42	-1.47	0.17	.00	[-1.79, -1.14]	[-1.74, -1.19]	-0.46
DimDown ^{FAT}	0.25	0.16	.12	[-0.06, 0.56]	[-0.01, 0.51]	0.07	-0.10	0.18	.58	[-0.46, 0.26]	[-0.41, 0.20]	-0.03
DimUp ^{FAT}	-0.67	0.19	.00	[-1.05, -0.29]	[-0.99, -0.35]	-0.22	-0.61	0.18	.00	[-0.96, -0.26]	[-0.91, -0.32]	-0.20
DimDown ^{VAT}	-1.06	0.18	.00	[-1.42, -0.71]	[-1.37, -0.76]	-0.36	-1.19	0.17	.00	[-1.52, -0.87]	[-1.47, -0.92]	-0.39
DimUp ^{VAT}	0.64	0.17	.00	[0.31, 0.98]	[0.37, 0.92]	0.20	0.63	0.17	.00	[0.29, 0.96]	[0.35, 0.90]	0.19
TemDown	-0.40	0.16	.01	[-0.71, -0.09]	[-0.66, -0.14]	-0.13	-0.33	0.16	.04	[-0.63, -0.02]	[-0.58, -0.07]	-0.10
TemUp	-0.09	0.16	.59	[-0.40, 0.23]	[-0.35, 0.18]	-0.03	-0.14	0.14	.35	[-0.42, 0.15]	[-0.37, 0.10]	-0.04
<i>Model Constraint</i>												
Social Comparison Effect	1.82	0.15	.00	[1.53, 2.11]	[1.58, 2.06]	0.62	2.09	0.17	.00	[1.77, 2.42]	[1.82, 2.36]	0.69
Dimensional Comparison Effect	0.92	0.19	.00	[0.55, 1.28]	[0.61, 1.22]	0.29	0.51	0.20	.01	[0.11, 0.91]	[0.17, 0.85]	0.17
Temporal Comparison Effect	-0.32	0.15	.04	[-0.62, -0.02]	[-0.57, -0.07]	-0.11	-0.19	0.15	.20	[-0.48, 0.10]	[-0.43, 0.05]	-0.06
Social - Dimensional Comparison Effect	0.90	0.24	.00	[0.44, 1.37]	[0.51, 1.30]	0.33	1.58	0.26	.00	[1.07, 2.10]	[1.15, 2.01]	0.52
Social - Temporal Comparison Effect	1.50	0.21	.00	[1.09, 1.92]	[1.15, 1.85]	0.51	1.90	0.22	.00	[1.47, 2.34]	[1.54, 2.27]	0.63
Dimensional - Temporal Comparison Effect	0.60	0.24	.01	[0.13, 1.07]	[0.20, 1.00]	0.18	0.32	0.25	.20	[-0.17, 0.81]	[-0.09, 0.73]	0.11
Social - Social Comparison Effect	-0.27	0.11	.02	[-0.49, -0.05]	[-0.46, -0.09]	-0.07						
Dimensional - Dimensional Comparison Effect	0.41	0.18	.02	[0.06, 0.76]	[0.11, 0.70]	0.12						
Temporal - Temporal Comparison Effect	0.13	0.12	.30	[-0.12, 0.37]	[-0.08, 0.33]	0.04						

FAT = Figural Ability Test. VAT = Verbal Ability Test. The specification of the effects under "Model Constraint" can be found in Fig. 4. N=297

Table 8 Results of the regression analyses predicting students' inferred VAT self-concept of a peer and student's own assessment of a peer's ability to solve VAT tasks (Model 4)

Predictor	Inferred VAT Self-Concept of a Peer				Assessment of a Peer's Ability to Solve VAT Tasks							
	Estimate	S.E	p	95% CI	90% CI	β	Estimate	S.E	p	95% CI	90% CI	β
Intercept	4.27	0.11	.00	[4.05, 4.48]	[4.09, 4.45]	4.93	4.35	0.11	.00	[4.12, 4.57]	[4.16, 4.54]	4.70
SocDown	0.55	0.14	.00	[0.26, 0.83]	[0.31, 0.78]	0.23	0.77	0.14	.00	[0.50, 1.04]	[0.55, 1.00]	0.30
SocUp	-1.14	0.17	.00	[-1.48, -0.81]	[-1.42, -0.86]	-0.41	-1.39	0.17	.00	[-1.73, -1.05]	[-1.68, -1.10]	-0.46
DimDown ^{FAT}	-1.12	0.20	.00	[-1.50, -0.74]	[-1.44, -0.80]	-0.32	-1.31	0.17	.00	[-1.64, -0.98]	[-1.59, -1.03]	-0.35
DimUp ^{FAT}	0.69	0.14	.00	[0.42, 0.96]	[0.46, 0.92]	0.27	0.86	0.13	.00	[0.61, 1.12]	[0.65, 1.08]	0.32
DimDown ^{VAT}	0.02	0.15	.90	[-0.27, 0.31]	[-0.23, 0.26]	0.01	0.06	0.15	.71	[-0.24, 0.35]	[-0.19, 0.30]	0.02
DimUp ^{VAT}	-0.38	0.18	.04	[-0.74, -0.02]	[-0.68, -0.08]	-0.12	-0.29	0.17	.10	[-0.63, 0.05]	[-0.57, -0.00]	-0.08
TemDown	-0.27	0.15	.07	[-0.56, 0.02]	[-0.52, -0.03]	-0.10	-0.20	0.15	.19	[-0.50, 0.10]	[-0.45, 0.05]	-0.07
TemUp	-0.23	0.16	.15	[-0.53, 0.08]	[-0.48, 0.03]	-0.08	-0.25	0.15	.11	[-0.55, 0.05]	[-0.50, 0.01]	-0.08
<i>Model Constraint</i>												
Social Comparison Effect	1.69	0.16	.00	[1.37, 2.00]	[1.42, 1.95]	0.63	2.16	0.15	.00	[1.86, 2.46]	[1.91, 2.41]	0.76
Dimensional Comparison Effect	0.40	0.18	.03	[0.05, 0.75]	[0.11, 0.69]	0.13	0.34	0.16	.03	[0.03, 0.66]	[0.08, 0.61]	0.10
Temporal Comparison Effect	-0.05	0.15	.77	[-0.34, 0.25]	[-0.29, 0.20]	-0.02	0.05	0.14	.73	[-0.23, 0.33]	[-0.19, 0.29]	0.01
Social - Dimensional Comparison Effect	1.29	0.24	.00	[0.82, 1.76]	[0.89, 1.68]	0.51	1.82	0.22	.00	[1.38, 2.25]	[1.45, 2.18]	0.66
Social - Temporal Comparison Effect	1.64	0.22	.00	[1.21, 2.07]	[1.28, 2.01]	0.61	2.11	0.21	.00	[1.70, 2.52]	[1.77, 2.46]	0.75
Dimensional - Temporal Comparison Effect	0.36	0.23	.13	[-0.10, 0.81]	[-0.03, 0.74]	0.10	0.29	0.22	.17	[-0.13, 0.72]	[-0.06, 0.65]	0.09
Social - Social Comparison Effect	-0.47	0.14	.00	[-0.75, -0.20]	[-0.70, -0.25]	-0.13						
Dimensional - Dimensional Comparison Effect	0.06	0.12	.65	[-0.19, 0.30]	[-0.15, 0.26]	0.02						
Temporal - Temporal Comparison Effect	-0.00	0.27	.99	[-0.53, 0.52]	[-0.45, 0.44]	-0.00						

FAT = Figural Ability Test. VAT = Verbal Ability Test. The specification of the effects under "Model Constraint" can be found in Fig. 5. N = 293

significantly stronger than the temporal comparison effect on students' inferred FAT self-concept of a peer ($B=0.60, p=.01, \beta=0.18$).

Three significant differences also emerged between the strength of the same comparison effects on either students' inferred self-concept of a peer or their own assessment of a peer's ability. First, the social comparison effect on students' inferred FAT self-concept of a peer was significantly weaker than the social comparison effect on students' own assessment of a peer's ability to solve FAT tasks ($B=-0.27, p=.02, \beta=-0.07$). Second, the social comparison effect on students' inferred VAT self-concept of a peer was significantly weaker than the social comparison effect on students' own assessment of a peer's ability to solve VAT tasks ($B=-0.47, p<.01, \beta=-0.13$). Third, the dimensional comparison effect on students' inferred FAT self-concept of a peer was significantly stronger than the dimensional comparison effect on students' own assessment of a peer's ability to solve FAT tasks ($B=0.41, p=.02, \beta=0.12$). No significant differences emerged between the strength of the dimensional comparison effects on students' inferred VAT self-concept of a peer and on their own assessment of a peer's ability to solve VAT tasks, or between the strength of the temporal comparison effects on students' inferred self-concept of a peer and on their own assessment of a peer's ability in the same domain (all $0.13 \leq |B|$, all $p \geq .30$).

An exploratory analysis of the effects of the dummy variables showed that for all four ability ratings examined in Model 3 and Model 4, students' ability ratings were significantly higher in the SocDown condition than in the Reference condition (all $0.52 \leq B \leq 0.77$, all $p < .01$, all $0.19 \leq \beta \leq 0.30$), whereas they were significantly lower in the SocUp condition than in the Reference condition (all $-1.47 \leq B \leq -1.14$, all $p < .001$, all $-0.46 \leq \beta \leq -0.41$). Students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks did not differ significantly between the Reference condition and the DimDown^{FAT} condition (all $0.25 \leq |B|$, all $p \geq .12$). In contrast, these ability ratings were significantly lower in the DimUp^{FAT} condition compared to the Reference condition (all $-0.67 \leq B \leq -0.61$, all $p < .01$, all $-0.22 \leq \beta \leq -0.20$). Similarly, students' inferred VAT self-concept of a peer and students' own assessment of a peer's ability to solve VAT tasks did not differ significantly between the Reference condition and the DimDown^{VAT} condition (all $0.06 \leq |B|$, all $p \geq .71$), but were significantly lower in the DimUp^{VAT} condition compared to the Reference condition (all $-0.38 \leq B \leq -0.29$, all $p \leq .10$, all $-0.12 \leq \beta \leq -0.08$). Students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks were significantly lower in the TemDown condition compared to the Reference condition (all $-0.40 \leq B \leq -0.33$, all $p \leq .04$, all $-0.13 \leq \beta \leq -0.10$). Students' inferred VAT self-concept of a peer and students' own assessment of a peer's ability to solve VAT tasks did not differ between the TemDown condition and the Reference condition (all $0.27 \leq |B|$, all $p \geq .07$). Similarly, students' inferred FAT self-concept of a peer and their own assessment of a peer's ability to solve FAT tasks, as well as students' inferred VAT self-concept of a peer and their own assessment of a peer's ability to solve VAT tasks, did not differ between the TemUp condition and the Reference condition (all $0.25 \leq |B|$, all $p \geq .11$).

6 Discussion

6.1 Main findings

The present study aimed to examine the joint effects of social, dimensional, and temporal comparisons on students' own self-concepts, students' inferred self-concepts of a peer, and students' own assessments of a peer's abilities. In accord with previous research, I found significant social and dimensional comparison effects on the different kinds of ability assessment. Furthermore, supporting TOESCI, all social comparison effects were significantly stronger than the other comparison effects.

The dimensional comparison effect on students' own FAT self-concept lost its significance after controlling for students' prior FAT self-concept. However, this finding can be explained by the fact that students in the condition triggering dimensional downward comparisons from the perspective of the FAT had already shown an increased FAT self-concept at the beginning of the study. Although this study does not provide clear evidence for the dimensional comparison effect on students' FAT self-concept, it is plausible that the manipulation to trigger the dimensional comparison effect on students' FAT self-concept showed no effect because of the pre-existing self-concept differences, which have limited the scope for further divergence in students' FAT self-concept. In any case, the absence of the dimensional comparison effect on students' FAT self-concept after controlling for their prior FAT self-concept should not be interpreted as evidence against the existence of dimensional comparison effects. This is especially true in light of the fact that all other dimensional comparison effects examined in this research were significant.

Despite the longitudinal design of the present study, the temporal comparison effects were mostly nonsignificant. This finding contradicts the findings of Wolff et al.'s (2018b) experimental study, Müller-Kalthoff et al.'s (2017a) vignette studies, and most non-experimental studies investigating the joint effects of social, dimensional, and temporal comparisons on students' self-concepts at school (e.g., Wolff & Möller, 2022). Still, it is consistent with the findings of the two experimental studies by Zell and Strickhouser (2020), and thus further calls into question the role of temporal comparisons in the formation of ability beliefs. Nevertheless, on the basis of this study's findings and those of Zell and Strickhouser (2020), it would be premature to conclude that temporal comparisons do not have an impact on the formation of students' ability beliefs. Rather, it is conceivable that the nonsignificant temporal comparison effects, as well as the significant temporal comparison effect on students' inferred FAT self-concept of a peer in the direction contrary to that expected, resulted from the fact that the experimental studies examined temporal comparisons between achievements in ability tests of similar difficulty. For this reason, participants could have interpreted achievement improvements to mean that students needed some practice in order to perform above-average, which could thus be indicate of relatively low domain-specific abilities. In contrast, achievement declines could have been considered as an indicator of relatively high domain-specific abilities, as students with achievement declines had demonstrated above-average achievement without much practice.

Unlike the experiments on the joint effects of social, dimensional, and temporal comparisons, temporal comparisons in the school context typically refer to achievements based on increasingly demanding assessment criteria. Thus, achievement improvements in school indicate that students have increased their competencies to an above-average extent, whereas achievement declines indicate that students have not increased their competencies to the desired extent. However, achievement declines in school do not imply that students' competencies have necessarily declined. Therefore, it is plausible that students in school are more inclined to interpret achievement improvements as indicative of high domain-specific abilities, compared to students who receive achievement feedback on two ability tests of similar difficulty. However, empirical research would be necessary to test this conjecture.

In addition to examining temporal comparison effects in a longitudinal experiment, a particular goal of this study was to examine whether the strength of comparison effects depends on the perspective from which students assess a peer's abilities. Interestingly, I found significantly stronger social comparison effects on students' own assessment of a peer's ability than on students' inferred self-concept of a peer in the figural and verbal domains, whereas the dimensional comparison effect on students' inferred FAT self-concept of a peer was significantly stronger than that on their own assessment of a peer's ability to solve FAT tasks. Hence, it seems that social comparisons have a particularly strong influence when students evaluate the abilities of others. In contrast, dimensional comparisons could be given greater weight when students (pretend to) assess their own abilities. A stronger focus on dimensional rather than social comparisons to assess one's own abilities would make sense, considering that the desire for self-differentiation has been shown to be a key reason for students to conduct dimensional comparisons (Wolff et al., 2018a). Nevertheless, this conclusion should be viewed with caution, since the dimensional comparison effects on students' inferred VAT self-concept of a peer and on their own assessment of a peer's ability to solve VAT tasks did not differ from each other.

6.2 Additional findings

As explained in Sect. 4.1, I operationalized the comparison effects by comparing students' ability ratings between those experimental conditions that triggered upward versus downward comparisons for the respective comparison type (and lateral comparisons for the other two comparison types). I used this approach to be consistent with previous studies and to achieve a high power. Nevertheless, analyses of the ability ratings in the conditions triggering upward and downward comparisons for one comparison type in relation to the ability ratings in the Reference condition, triggering lateral comparisons for all three comparison types, yielded some interesting findings.

The first interesting finding is that five of six ability ratings (students' own VAT self-concept, students' inferred FAT and VAT self-concepts of a peer, and students' own assessments of a peer's ability to solve FAT and VAT tasks) were significantly lower in the corresponding conditions triggering dimensional upward comparisons compared to the Reference condition, whereas none of the ability ratings

differed significantly between the corresponding conditions triggering dimensional downward comparisons and the Reference condition. This finding contradicts the assumption of a positive net effect of dimensional comparisons, originally formulated in dimensional comparison theory (Möller & Marsh, 2013). According to this effect, the positive effects of dimensional downward comparisons on students' self-concepts would be stronger than the negative effects of dimensional upward comparisons. However, although initial studies suggested the existence of a positive net effect of dimensional comparisons (Pohlmann & Möller, 2009), a more recent meta-analysis did not find evidence of a positive net effect (Müller-Kalthoff et al., 2017b). The finding of the present study that for most ability ratings examined, only the effects of dimensional upward comparisons were significant, provides additional support for this conclusion. Moreover, one could even be inclined to interpret this finding as evidence of a negative net effect of dimensional comparisons. However, it is important to note that the effects of dimensional upward and downward comparisons did not differ significantly for any of the six ability ratings (as can be inferred from the 95% confidence intervals), so that this conclusion would go too far.

A second finding worth discussing is that three of the four external ability ratings (students' inferred FAT self-concept of a peer and students' own assessments of a peer's ability to solve FAT and VAT tasks) were significantly more strongly affected by the feedback triggering social upward comparisons than by the feedback triggering social downward comparisons (as can also be inferred from the 95% confidence intervals). Thus, the students in this study were more inclined to devalue the abilities of their peers performing below-average than to enhance the abilities of their peers performing above-average. An explanation for this finding could be self-enhancement motivations, since the students could have been able to develop a more positive perception of their own abilities by devaluing their peers' abilities (e.g., Wolff et al., 2018a). This explanation would also align with the finding that the effects of social upward and downward comparisons did not differ when students assessed their own self-concepts (although stronger effects of social downward comparisons would also have been plausible to serve self-enhancement motivations). The finding that stronger effects of social upward comparisons than of social downward comparisons were also shown on students' inferred FAT self-concept of a peer seems to contradict the assumption that the stronger effects of social upward comparisons resulted from self-enhancement motivations, at least at first glance. However, it is conceivable that self-enhancement motivations had such a strong (unconscious) influence in this study that the students were only partially successful in putting themselves into the shoes of their peers, to infer their self-concepts.

Finally, it is interesting to note that students' inferred FAT self-concept of a peer and students' own assessment of a peer's ability to solve FAT tasks were significantly lower in the condition triggering temporal downward comparisons compared to the Reference condition, although temporal downward comparisons were assumed to enhance students' ability ratings. Moreover, in all other analyses, students' ability ratings in the conditions triggering temporal downward comparisons were descriptively lower than in the Reference condition. Given these findings, one could speculate that students tend to rate abilities more highly in the case of constant

achievement across time, than in the case of achievement changes (around the same achievement level). Although this assumption is plausible, especially in regard to developments in achievement in tasks of equal difficulty (as was the case in the present study; see also discussion about temporal comparison effects in Sect. 6.1), it would be interesting to pursue this hypothesis in more detail in future research. In particular, this is true for non-experimental studies, in which researchers have usually investigated the effects of students' achievement changes on their self-concepts using linear rather than quadratic regressions.

6.3 Strengths and limitations

This study has a number of strengths. In particular, it has examined the joint effects of social, dimensional, and temporal comparisons, for the first time, in a longitudinal experiment, for different types of ability assessment, and in different domains. Nevertheless, the study also has some limitations.

First, in many cases the experimental manipulation was not successful. Perhaps this was because the achievement feedback, consisting of four percentile ranks, included too much information for some students. This explanation seems plausible, given that the students were particularly likely to respond to the items that asked them to rate their own or their peer's achievement in dimensional comparison (i.e., those items that required the students to relate all four percentile ranks to each other, rather than just two percentile ranks) in a way that was inconsistent with the experimental manipulation. Because of manipulation check failures, I had to exclude a significant number of students from the analyses. Moreover, I conducted the analyses with different subgroups of students, to avoid having to exclude more participants. Nonetheless, even after excluding students, this study still had acceptable power and more participants than most previous experimental studies examining comparison effects.

Second, it is possible that realizing this study in small groups and integrating the feedback about a peer's achievement triggered the salience of social comparisons. Consequently, students may have paid relatively little attention to the dimensional and temporal comparison information when making their ability assessments. However, the design of this study has the advantage of a relatively high ecological validity. For example, students at school usually take achievement tests in groups (i.e., within their class) and often receive information not only about their own achievement, but also about the achievement of individual classmates (e.g., their seatmate), when their teachers give them achievement feedback.

Finally, it is worth mentioning that the students did not receive feedback on their achievement in the ability tests from T1 before T2. For ethical reasons, I refrained from leaving students alone with manipulated feedback for two weeks. Nevertheless, it is possible that temporal comparisons would have had a (stronger) impact on the formation of students' ability beliefs if students had already received achievement feedback at T1.

7 Conclusion

This research provides additional support for the causal impact of social and dimensional comparisons on different types of ability belief. Yet, it remains unclear to what extent temporal comparisons are involved in the formation of students' ability beliefs. My assumption that previous experimental studies did not find significant temporal comparison effects because the temporal comparisons related to periods of only a few minutes no longer seems plausible in view of the results of this study. Still, further research is necessary to explain the inconsistent findings on the influence of temporal comparisons on students' ability assessments from different (kinds of) studies. This is particularly important with respect to the development of a comprehensive comparison theory of academic self-concept formation that integrates social, dimensional, and temporal comparisons.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11218-023-09808-9>.

Acknowledgements I thank Ole Gries, Fynn Junge, Nina Kaufmann, Jasmin Lehrmann, Alexander Litwinov, Marie-Claire Menikheim, Lea Michl, and Alexander Reisinger for their help with data collection. I also thank Stephen McLaren for copyediting the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was funded by the German Research Foundation under Grants WO 2341/1-1 and WO 2341/2-1.

Data availability The data that support the findings of this study can be downloaded at https://osf.io/xt8vh/?view_only=a0020496aafb47d4bbe11c3e0c88cd7. The syntaxes used to analyze the data can be found in the Supplemental Material.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

Ethical approval The local ethics committee of Kiel University had approved the study prior to data collection (Approval no.: D 537/18). The participants gave informed consent to participate in the study at the beginning and at the end of the study (after being informed about the actual purpose of the study).

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(6), 485–503. <https://doi.org/10.1037/0033-295X.84.6.485>

- Dai, D. Y. (2002). Incorporating parent perceptions: A replication and extension study of the internal-external frame of reference model of self-concept development. *Journal of Adolescent Research, 17*(6), 617–645. <https://doi.org/10.1177/074355802237467>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. <https://doi.org/10.1177/001872675400700202>
- Helm, F., Müller-Kalthoff, H., Mukowski, R., & Möller, J. (2018). Teacher judgment accuracy regarding students' self-concepts: Affected by social and dimensional comparisons? *Learning and Instruction, 55*, 1–12. <https://doi.org/10.1016/j.learninstruc.2018.02.002>
- Helm, F., Wolff, F., Möller, J., Zitzmann, S., Marsh, H. W., & Dicke, T. (2023). Individualized teacher frame of reference and student self-concept within and between school subjects. *Journal of Educational Psychology, 115*(2), 309–329. <https://doi.org/10.1037/edu0000737>
- Jansen, M., Becker, M., & Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *Journal of Educational Psychology, 113*(2), 330–350. <https://doi.org/10.1037/edu0000524>
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)* [Intelligence Structure Test 2000R (I-S-T 2000 R)] (2nd ed.). Hogrefe.
- Lösch, T., Kelava, A., Nagengast, B., Trautwein, U., & Lüdtke, O. (2017). Perspective matters: The internal/external frame of reference model for self- and peer ratings of achievement. *Learning and Instruction, 52*, 80–89. <https://doi.org/10.1016/j.learninstruc.2017.05.001>
- Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction, 33*, 50–66. <https://doi.org/10.1016/j.learninstruc.2014.04.002>
- Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. *Australian Journal of Psychology, 36*(3), 367–386. <https://doi.org/10.1080/00049538408255318>
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*(2), 397–416. <https://doi.org/10.1111/j.1467-8624.2005.00853.x>
- Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology, 93*(4), 826–835. <https://doi.org/10.1037/0022-0663.93.4.826>
- Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review, 120*(3), 544–560. <https://doi.org/10.1037/a0032459>
- Müller-Kalthoff, H., Helm, F., & Möller, J. (2017a). The Big Three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Social Psychology of Education, 20*(4), 849–873. <https://doi.org/10.1007/s11218-017-9395-9>
- Müller-Kalthoff, H., Jansen, M., Schiefer, I. M., Helm, F., Nagy, N., & Möller, J. (2017b). A double-edged sword? On the benefit, detriment, and net effect of dimensional comparison on self-concept. *Journal of Educational Psychology, 109*(7), 1029–1047. <https://doi.org/10.1037/edu0000171>
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Muthén & Muthén.
- Pohlmann, B., & Möller, J. (2006). Vergleichseffekte auf kognitive, affektive und motivationale Variablen: Eine experimentelle Überprüfung [Comparison effects on cognitive, affective, and motivational variables: An experimental approach]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(2), 79–87. <https://doi.org/10.1026/0049-8637.38.2.79>
- Pohlmann, B., & Möller, J. (2009). On the benefit of dimensional comparisons. *Journal of Educational Psychology, 101*(1), 248–258. <https://doi.org/10.1037/a0013151>
- Pohlmann, B., Möller, J., & Streblov, L. (2004). Zur Fremdeinschätzung von Schülerelbstkonzepten durch Lehrer und Mitschüler [On students' self-concepts inferred by teachers and classmates]. *Zeitschrift für Pädagogische Psychologie, 18*(3/4), 157–169. <https://doi.org/10.1024/1010-0652.18.34.157>
- 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>
- 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*(5), 676–688. <https://doi.org/10.1177/0146167214523475>
- Van Zanden, B., Marsh, H. W., Seaton, M., Parker, P. D., Guo, J., & Duineveld, J. J. (2017). How well do parents know their adolescent children? Parent inferences of student self-concepts reflect

- dimensional comparison processes. *Learning and Instruction*, 47, 25–32. <https://doi.org/10.1016/j.learninstruc.2016.09.002>
- Wolff, F., Helm, F., Junge, F., & Möller, J. (2020a). Are dimensional comparisons performed unconsciously? An investigation of the internal/external frame of reference model using implicit self-concepts. *Journal of Educational Psychology*, 112(2), 397–415. <https://doi.org/10.1037/edu0000375>
- 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(4), 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(7), 1005–1025. <https://doi.org/10.1037/edu0000248>
- Wolff, F., & Möller, J. (2022). An individual participant data meta-analysis of the joint effects of social, dimensional, and temporal comparisons on students' academic self-concepts. *Educational Psychology Review*, 34(4), 2569–2608. <https://doi.org/10.1007/s10648-022-09686-1>
- Wolff, F., Nagy, G., Retelsdorf, J., Helm, F., Köller, O., & Möller, J. (2019). The 2I/E model: Integrating temporal comparisons into the internal/external frame of reference model. *Journal of Educational Psychology*, 111(7), 1131–1161. <https://doi.org/10.1037/edu0000319>
- Wolff, F., Sticca, F., Niepel, C., Götz, T., Van Damme, J., & Möller, J. (2021). 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(8), 1529–1549. <https://doi.org/10.1037/edu0000632>
- Wolff, F., Wigfield, A., Möller, J., Dicke, T., & Eccles, J. S. (2020b). 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(8), 1644–1660. <https://doi.org/10.1037/edu0000440>
- Zell, E., & Alicke, M. D. (2009). Self-evaluative effects of temporal and social comparison. *Journal of Experimental Social Psychology*, 45(1), 223–227. <https://doi.org/10.1016/j.jesp.2008.09.007>
- Zell, E., & Alicke, M. D. (2010). Comparisons over time: Temporal trajectories, social comparison, and self-evaluation. *European Journal of Social Psychology*, 40(3), 375–382. <https://doi.org/10.1002/ejsp.737>
- Zell, E., & Strickhouser, J. E. (2020). Comparisons across dimensions, people, and time: On the primacy of social comparison in self-evaluations. *Social Psychological and Personality Science*, 11(6), 791–800. <https://doi.org/10.1177/1948550619884564>

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

Fabian Wolff is Junior Professor in Educational Psychology and Empirical Research Methods and head of the Emmy Noether Junior Research Group COMPASS (“Comparison Processes in Students’ Academic Self-Concept Formation”) at the Institute of Psychology at the University of Koblenz. In his dissertation and postdoctoral (habilitation) thesis, he focused on the impact of comparisons in the process of academic self-concept formation. Besides the formation of academic self-concepts, he is particularly interested in the development of competencies and the digitization of learning environments.